Agricultural Economics Research

Summer 1984 Vol. 36 No. 3

A Journal of the United States Department of Agriculture

Economic Research Service

Agricultural Economics Research

A Journal of the U.S. Department of Agriculture • Economic Research Service

The Secretary of Agriculture has determined that the publication of this periodical is necessary in the transaction of public business required by law of this Department. Use of funds for printing this periodical has been approved by the Director, Office of Management and Budget, through March 31, 1987.

This periodical is published quarterly. Subscription rates are $11.00 domestic; $13.75 foreign. Send check or money order to (payable to Superintendent of Documents):

Superintendent of Documents
U.S. Government Printing Office
Washington, D.C. 20402

Microfiche copies are available from two different sources:

Infordata International, Inc.
Suite 4602
175 East Delaware Place
Chicago, Ill. 60611

Issues are available from 1974 on at $4.95 per year or per issue.

Congressional Information Service, Inc.
P.O. Box 30056
Washington, D.C. 20014

Issues for 1973 are available at $5.00 (for 2 fiche). Issues for 1974 on are available at $3.75 for the first fiche and $1.25 for each remaining fiche; they average 3 fiche per year.

In This Issue

This issue emphasizes the need to examine underlying assumptions in economic analysis. The research article examines the underlying logical and statistical assumptions of a research technique which has experienced a wave of popularity in economic journals. Two short papers in the Research Review section apply two widely differing techniques: experimentation with an assumed cost function and long-term time trend analysis which emphasizes the importance of analysts examining their assumptions.

Conway, Swamy, Yanagida, and von zur Muehlen examine causality tests as developed by Sims and Granger. These tests examine time series for evidence that one time series "causes" another time series. Numerous articles using these tests have appeared in the economic literature; at least four have appeared in this journal. Conway, Swamy, Yanagida, and von zur Muehlen's article will be a watershed article in this body of economic literature. Their findings, "Causality tests developed by Sims and Granger are fatally flawed," cannot be ignored. Future users of these causality tests will find it necessary to weigh the arguments in their article.

The Research Review section continues the theme of examining underlying assumptions in economic analysis. Lutton contends that the analyst's view of input substitution potential in future agricultural production is a crucial assumption in the debate on worldwide agricultural capacity. Pessimists implicitly assume low input substitution potential, and impending resource constraints feed their pessimism. Optimists see a more flexible production system. Lutton uses an assumed cost function and alternative elasticity-of-substitution estimates to demonstrate that an analyst's view of the input substitution potential is not inconsequential to the debate.

Edwards considers another aspect of the debate about the future real cost of food. He examines the price history of a commodity during most of this century to gain insight into the future real cost of food. He concludes that the longrun trend of real food prices is downward and volatile, and the burden of proof is on those who expect otherwise.

Boxley concludes this section with a review of Marion Clawson's *The Federal Lands Revisited*. This book benefits from longrun analysis of another form as Clawson has been professionally concerned with the Federal lands of the United States for 45 years.

Gerald Schluter

Contents

The Impossibility of Causality Testing

By Roger K. Conway, P. A. V. B. Swamy, John F. Yanagida, and Peter von zur Muehlen*

Abstract

Causality tests developed by Sims and Granger are fatally flawed for several reasons. First, when two variables, X and Y, are uncorrelated, X has no *linear* predictive value for Y; but X and Y may be nonlinearly related unless they are statistically independent, in which case X and Y are not related at all. The right-hand side variables in a regression equation are exogenous if they are mean independent of the disturbance term. Mean independence is stronger than uncorrelatedness. The proofs for deriving causality-exogenity tests imply weaker results than statistical or mean independence. Second, transformations such as the Box-Cox transformation and Box-Jenkins stationarity-inducing transformations are not causality preserving. Third, counterexamples constructed by Price have invalidated the Pierce-Haugh theorem on instantaneous causality. Fourth, omission of other variables influencing those tested renders any test results spurious. Finally, causality tests are inconsistent because they are based on underidentified models. We provide a logically valid method of building models which does not use causality tests.

Keywords

Causality tests, statistical independence, mean independence, uncorrelatedness, orthogonality, covariance stationarity, stationarity-inducing transformations, economic laws

"Neglect by theorists evokes malpractice by empiricists."

Arthur S. Goldberger (*30*)[1]

Introduction

Numerous recent studies in the agricultural literature use or proselytize tests of causality originally developed by Sims (*58*).[2] The theoretical basis of this test is reproduced in Sargent (*54*, pp. 285-87). (For further discussion, see (*52*).) In an earlier study, Sargent (*53*) describes a causality test procedure, attributable to Granger (*26*) which is different from Sims' procedure. Both of these tests employ the following Granger (*26*) concept of causality: A time series (x_t) Granger causes another time series (y_t) if one can predict present y better by using past values of x than by not doing so. For example, in a given bivariate covariance stationary stochastic process (y_t, x_t) possessing a vector autoregressive representation, y fails to Granger cause x if and only if the coefficient matrices of the process are upper triangular. (We use the term, "Granger cause," to refer to causality in Granger's sense.) This result holds because the upper triangularity of coefficient matrices implies that y_t

*Conway is an economist with ERS; Swamy and von zur Muehlen are senior economists at the Federal Reserve Board, Washington, D.C., and Yanagida is an associate professor of agricultural economics at the University of Nevada at Reno. Views in this article are the authors' and do not reflect those of the Federal Reserve Board or the U.S. Department of Agriculture. The authors received valuable comments and help from Lorna Aldrich, James Barth, Michael Bradley, Richard Haidacher, Charles Hallahan, Arthur Havenner, Anil Kashyap, Nadine Loften, Thomas Lutton, Lloyd Teigen, Michael Weiss, and especially J. Michael Price. The authors are also grateful to David A. Pierce whose remarks are incorporated into this article.

[1] Italicized numbers in parentheses refer to items in the References at the end of this article.

[2] For example, see (*4, 5, 6, 8, 34, 39, 60, 67*).

can be expressed as a distributed lag of current and past x's (with no future x's) with a disturbance process denoted by u_t and that past y's do not help predict x_t, given past x's. However, the disturbance u_t is uncorrelated with past, present, and future x's for only one value of ρ in the regression $a_{yt} = \rho a_{xt} + \xi_t^{\rho}$ where (a_{yt}, a_{xt}) is the vector of innovations of the process (y_t, x_t). If the coefficient matrices are not triangular, then u_t is not uncorrelated with past, present, and future x's for any value of ρ. Because the value of ρ is usually unknown, for the disturbance u_t in the regression of y_t on current and past x's to be uncorrelated with past, present, and future x's, it is necessary, but not sufficient, that y fails to Granger cause x or that the coefficient matrices of the process (y_t, x_t) are upper triangular. The null hypothesis for Granger's causality test is that the coefficient matrices of the process (y_t, x_t) are upper triangular. This hypothesis can be equivalent to Sims' hypothesis that all the coefficients of future x's in the regression of y on past, present, and future x's are zero. Thus, Sims and Granger try to test a necessary condition for Granger noncausality. These tests were formerly associated predominantly with research in macroeconomics, which tests monetarist *versus* Keynesian assumptions about the causal ordering between money and income. They have recently been used in conjunction with rational expectations hypothesis testing.[3] Various studies using the same testing procedures have produced contradictory evidence on the relationship between money and income (see (*59*)). The conflict between the conclusions of such studies were indeed heightened when different forms of causality testing procedures were employed (see (*21*)). Subsequent Monte Carlo tests offered suggestive results indicating differences in the power of various causality tests and showing that one could easily produce conflicting conclusions by employing a battery of causality tests on the same data sets (see (*25*, *38*)). However, these empirical and Monte Carlo results are only symptomatic. It is now clear that there are profound problems, both theoretical and empirical, with causality tests. This viewpoint is most emphatically stated by statisticians who object to the apparent carelessness with which some economists equate correlation with causality (see (*37*)). The purpose of our article is, therefore, to alert the agricultural profession to these problems and to allow agricultural researchers to better weigh the benefits and costs of utilizing these tests.

With that purpose in mind, we establish the following points:

1. The zero correlation between u_t and past, present, and future x's is necessary, but not sufficient, for x to be strictly econometrically exogenous with respect to y. The proofs of causality and exogeneity advanced by proponents are based on weaker concepts than statistical or mean independence.

2. There is no good discriminant between stationary and nonstationary processes. Sims and Granger are testing a necessary condition for Granger noncausality only within the framework of covariance stationary processes.

3. The observed time series is necessarily finite, and the covariance stationary stochastic processes are infinite in length. Distinguishing between different stationary processes on the basis of observed time series poses fundamental difficulty. Therefore, the power of Sims' or Granger's test does not go to 1 as the sample size goes to infinity.

4. Even if we know the transformations which induce stationarity, these transformations are not causality preserving. Therefore, the causality relationships (or the lack thereof) among the transformed variables tell us nothing about the causality relationships (or the lack thereof) among the original variables.

5. Zellner (*70*) proposed a general definition of causality attributed to Feigl, according to whom the concept of causation is defined in terms of predictability according to a law. Therefore, we address a fundamental question in economics: Are there laws in economics? After answering this question, we suggest a logically valid method of building econometric models which does not use causality tests.

[3] A key requirement of rational expectations observable reduced-form equations is that all right-hand side variables be at least orthogonal to the error term (see (*17*, *18*)).

In subsequent sections, we define the various notions of Granger causality and contrast them with what statisticians call statistical or mean independence. We discuss problems of forming conditional operations based on linear models. We describe and critique the characterizations of Granger causality noted by Pierce and Haugh (*41*). We consider the causality tests as Sims proposed. We offer some general remarks on causality testing. Because we refer to laws in a philosophical definition of causation, we briefly discuss the meaning of the term "law" in economic contexts.

Correct Interpretations of Granger's Definitions of Causality

Before the causality literature can be carefully critiqued, we need to understand exactly what is meant by "causality" as posited by its proponents. Therefore, we review the various forms of Granger causality defined by Granger (*26*) and extended by Pierce and Haugh (*41*). A_t is assumed to represent a stationary stochastic vector process where:

$\bar{A}_t$ = the set of past values of A_t;

$\bar{\bar{A}}_t$ = the set of past and present values of A_t;

$\bar{A}_t(k)$ = the set $(A_{t-j}, j \geqslant k)$;

$E_t(A|B)$ = the optimal predictor of A_t, given some set of values of B_t;[4]

$e_t(A|B)$ = the prediction error = $A_t - E_t(A|B)$;

$Var(e_t)$ = $\sigma^2\ (A|B)$;

U_t = the set of all information in the universe accumulated since time t-1; and

$U_t - Y_t$ = all information in the universe apart from Y_t.

With this information, we can define the various forms of causality as follows:

[4] By use of a mean square error or quadratic loss criterion.

1. Causality: If $\sigma^2(X|\bar{U}) < \sigma^2\ (X|\bar{U} - \bar{Y})$, then we say Y is Granger causing X, denoted by $Y_t \Rightarrow X_t$.

2. Feedback: There is feedback between X and Y, denoted by $X_t \Leftrightarrow Y_t$, if $Y_t \Rightarrow X_t$ and if $X_t \Rightarrow Y_t$.

3. Instantaneous causality: Instantaneous causality occurs when $\sigma^2\ (X|\bar{U}, \bar{\bar{Y}}) < \sigma^2\ (X|\bar{U})$.

4. Causality lag: If $Y_t \Rightarrow X_t$, we then define the causality lag m as the lowest integer value of k so that the $\sigma^2(X|U - Y(k)) < \sigma^2(X|U - Y(k+1))$.

We now show that these definitions cannot be used to discover causality relationships without their posing some serious problems. Specifically, Granger's definitions require unequal and finite mean square errors in the series being compared. These conditions may not be satisfied in practice as may be clarified if one considers two simple polar cases. Deterministic variables or components can be predicted perfectly by their own past history with zero mean square error (see (*2*, p. 420)); hence, the mean square errors of the predictions of deterministic components do not satisfy the strict inequalities stated in Granger's definitions. This limitation, however, does not mean that there are no causality relationships among deterministic components. At the other extreme, when the mean square errors of the predictions of stochastic variables are infinite (a frequent occurrence in practice), Granger's definitions stated in terms of the strict inequalities between finite mean square errors of predictions do not apply. The fundamental problems associated with Granger's definitions will be clearer once we discuss the statistician's definitions and interpretations of statistical independence, mean independence, uncorrelatedness, and orthogonality.[5]

The variable Y is said to be statistically independent of the variable X if the conditional distribution of

[5] Related to this discussion are three recent papers by Chamberlain (*15*), Florens and Mouchart (*22*), and Engle, Hendry, and Richard (*19*) also expressing certain limitations of Granger's and Sims' definitions of causality. We extend their work by explicitly contrasting various notions of Granger causality with the statistician's concept of statistical independence or mean independence.

Y, given X = x, is the same as the marginal distribution of Y; that is, $F(y|x) = F(y)$, in which case:

$$F(y,x) = F(y)\,F(x) \tag{1}$$

where $F(y,x)$ is the joint distribution of Y and X, and $F(x)$ and $F(y)$ are the marginal distributions of X and Y, respectively. Then the conditional distribution of X, given Y = y, denoted by $F(x|y)$, is equal to $F(x)$; that is, X is independent of Y. These two variables, Y and X, are said to be independent if equation (1) holds, including the case where $F(y)$ or $F(x)$ is zero. It is difficult to establish the existence of $F(y|x)$ or $F(x|y)$ in the general case. The conditional probability of a set $A \in B$ (a Borel field of sets), given X = x, can be exhibited as a conditional expectation if one chooses the random variable Y as the indicator function of the set A. Thus, $P(A|x) = E(Y|x)$, as may be verified from the definition of conditional probability as given by Rao (*48*, p. 90), for example. One should note that the Radon-Nikodym theorem establishes the existence of $P(A|x)$ almost everywhere with respect to $[dF(x)]$ as a function of x for fixed A only where the exceptional x-set may depend on A. As a result, it may not be possible to define $P(A|x)$ for all A over an x-set of probability 1, unless the union of exceptional sets is of probability zero. Thus, a conditional probability distribution of Y, given X = x, may not always exist (see (*48*, p. 98)).[6] The same is true of the conditional probability distribution of X, given Y = y. Because the existence of $F(y|x)$ does not imply the existence of $F(x|y)$, if $F(y|x) = F(y)$, it need not be true that $F(x|y) = F(x)$. Nonetheless, when equation (1) is true, X and Y are said to be independent regardless of whether $F(y|x)$ or $F(x|y)$ exists.

The intuitive idea of the phrase "Y is independent of X" is roughly that a knowledge of X does not help one to infer the value of Y. If Y and X are statistically independent, then there is no causal relationship between Y and X. When $F(\cdot)$ and $F(\cdot, \cdot)$ are absolutely continuous, the probability density functions exist and equation (1) can be expressed as:

$$f(y,x) = f(y)f(x) \tag{2}$$

where $f(\cdot)$ is a density function.

As Whittle (*68*, p. 101) points out, we must live with the idea that we may know $E(Y)$ (or $E(X)$) only for certain Y (or X), or that, for a given random variable Y (or X), we may know $EK(Y)$ (or $EH(X)$) only for certain K (or H). Similarly, for a given pair of random variables, Y and X, we may be able to assert the validity of the independence condition:

$$EH(X)K(Y) = EH(X)EK(Y) \tag{3}$$

where the functions H and K are such that $EH(X) < \infty$ and $EK(Y) < \infty$. In this case, Y and X have only partial degrees of independence because equation (1) implies equation (3), but the converse is not true. An extreme example of this is one where we can assert the validity of the independence condition (equation (3)) only when H and K are linear functions. This essentially means we know only that:

$$EXY = EXEY \tag{4}$$

where $EX < \infty$ and $EY < \infty$. Two random variables, X and Y, are said to be uncorrelated if and only if both have finite second moments and equation (4) is true (see (*16*, p. 102)). Consequently, equation (4) is equivalent to:

$$\text{Cov}\,(X,Y) = 0 \tag{5}$$

provided $EX^2 < \infty$ and $EY^2 < \infty$. Random variables that satisfy equation (5) are said to be uncorrelated. In the special case when either $EX = 0$ or $EY = 0$, so that equation (4) becomes $EXY = 0$, the random variables are said to be mutually orthogonal. According to Whittle (*68*, p. 102), "the concept of lack of correlation or orthogonality is important, because it is the nearest one can come to the concept of independence if one is restricted to a knowledge of second moments [as in the case of covariance stationary processes]."

---

[6] If the sample space has only a countable number of points, then the conditional probability measure is always defined, provided $P(X = x) \neq 0$. Alternatively, if the sample space is the n-dimensional real Euclidean space, then the conditional probability measure exists because in this case the union of exceptional sets is of zero probability measure (*48*, pp. 98-99). Our subsequent discussion further clarifies this point.

Just as independence means that X has no predictive value for Y, lack of correlation means that X has no predictive value for Y *in the linear least squares sense* (see (*68*, p. 102)). That is, suppose we consider a predictor of Y which is linear in X, $Y = \alpha + \beta X + U$, and we choose α and β so as to minimize EU^2. One can then determine the optimal value of β by Cov(Y,X)/ Var(X). Thus, if case (5) is true, the variable X will receive a zero coefficient in the prediction formula for Y. When case (5) is true, X has no *linear* predictive value for Y, but X may be nonlinearly related unless equation (1) is true, in which case X and Y are not related at all.

A case intermediate between lack of correlation and independence is that in which equation (3) holds only for linear K, so that $EH(X)Y = EYEH(X)$ for any H assuming $EH(X) < \infty$. The relation $EH(X)Y = EYEH(X)$ is equivalent to $E(Y|x) = EY$ because $EH(X)Y = E(E[H(X)Y|X]) = E[H(X)E(Y|X)]$ for all H so that $EH(X)Y < \infty$ (see (*68*, p. 102)).[7] Here $E(Y|x)$ is a function of x, say G(x), which minimizes $E[Y - G(x)]^2$, at least in the case where $EY^2 < \infty$ (see (*2*, pp. 417-24)). Following Goldberger (*31*), we may say that Y is *mean independent* of X if:

$$E(Y|x) = EY \tag{6}$$

Now equation (6) holds if and only if $E(Ye^{itX}) = EYEe^{itX}$ for all real t (see (*36*, p. 10)).

It is instructive to observe that without further conditions there is no connection among the concepts (1), (5), and (6). If EY exists, it follows from the Radon-Nikodym theorem that $E(Y|x)$ exists (see Rao (*48*, pp. 96-97)). In this case, equation (1) implies equation (6), but the converse is not true. Similarly, if EX exists, then equation (1) implies the condition, $E(X|y) = EX$, but the converse is not true. Because the existence of EH(X) and EK(Y) is already assumed in condition (3), partial independence condition (3) implies the mean independence condition, $E(Y|x) = EY$ or $E(X|y) = EX$, but the converse is not true. It is obvious that any pair of random variables, X and Y, which are fully independent in the sense of equation (1) and which have finite variances are also uncorrelated, although the converse is not true. When X and Y have finite variances, mean independence (6) implies uncorrelatedness (5), but the converse is not true. (In the normal case, conditions (1-6) are equivalent.)

Our discussion is important as, when Granger's definitions of causality are used, some researchers have confused these statistical concepts. For example, Sargent (*52*, pp. 404-05) says that X in the following equation:

$$Y_t = \Sigma_{j=0}^{\infty} h_j X_{t-j} + U_t \tag{7}$$

with $\Sigma_{j=0}^{\infty} |h_j| < \infty$, $EU_t = 0$, $EU_t^2 = \sigma_u^2$ for all t, and $EU_tU_s = 0$ for $t \neq s$, is econometrically exogenous with respect to Y_t if and only if $EU_tX_s = 0$ for *all* integers s and t. This definition runs counter to some textbook notions of exogeneity. For example, Theil (*65*, pp. 110-11) and Goldberger (*29*, pp. 380-81) have stated that X in equation (7) is econometrically exogenous with respect to Y if $E(U_t|X_s) = EU_t = 0$ for *all* integers s and t. This condition is stronger than Sargent's condition, as shown by the direction of the implication between equations (5) and (6).[8] Furthermore, in his statement about a stricter form of the natural rate hypothesis, Sargent (*53*, p. 215) incorrectly equates condition (1) with condition (6) by saying that the unemployment rate Un_t obeys the natural rate hypothesis if, in its univariate Wold representation (without a purely deterministic component):

$$Un_t = \Sigma_{j=0}^{\infty} a_j U_{t-j}, \ \Sigma_{j=0}^{\infty} |a_j| < \infty \tag{8}$$

where the U's are serially uncorrelated with mean zero and finite variance, σ_u^2, the innovation U_t satisfies the condition:

$$E(U_t|\theta_{t-1}) = 0 \tag{9}$$

where θ_t is a vector of the set of all variables observed at time t thought potentially to contribute to predicting unemployment, so that the innovation in the unemployment rate is statistically independent of each component of θ_{t-1}. Here some elements of θ_t represent policy instruments. Another difficulty is that Sargent's time series methods

[7] One should note that when further expectation is taken, $E(Y|x) = E(Y|X = x)$ is replaced by $E(Y|X)$ (see (*48*, p. 97)).

[8] The direction of this implication has been recognized only recently by Hayashi and Sims (*33*).

based on non-Gaussian assumptions are only capable of examining the validity of the uncorrelatedness assumption between U_t and an element of θ_{t-1} but not the validity of the mean independence assumption (9) between these two variables.

Conditional Expectations and Econometric Modeling

Note that the existence of $E(Y|x)$ does not imply the existence of $E(X|y)$.[9] Necessary and sufficient conditions for the existence of the linear population regression function, $E(Y|x) = \alpha + \beta x$, and the constant conditional variance, $\mathrm{Var}(Y|x) = \sigma_0^2$, have been established by Rao (see (*36*, p. 11, lemma 1.1.3)). Generalized conditions covering the cases of several independent variables are given by Kagan, Linnik, and Rao (*36*), hereafter referred to as KLR. Because these conditions have far-reaching implications for causality tests, we state them here:

KLR's lemma (*36*): Let $\phi(t_0, t_2,..., t_K)$ be the characteristic function of the vector variable $(Y_t^*, X_{2t}^*,..., X_{Kt}^*) = (Y_t, X_{2t},..., X_{Kt}) - E(Y_t, X_{2t},..., X_{Kt})$. Then, for the relations $E(Y_t|x_{1t},..., x_{Kt}) = \sum_{k=1}^{K} \pi_k x_{kt}$ with $x_{1t} = 1$ and $\mathrm{Var}(Y_t|x_{1t},..., x_{Kt}) = \sigma_0^2$ = a positive constant $(t = 1, 2,..., T)$ to hold, it is necessary and sufficient that for $t = 1, 2,..., T$:

$$D_0\phi(t_0, t_2,..., t_K)\,|_{t_0=0} = \sum_{k=2}^{K} \pi_k D_k\phi(0, t_2,..., t_K),$$

$$D_0^2\phi(t_0, t_2,..., t_K)|_{t_0=0} = -\sigma_0^2\phi(0, t_2,..., t_K) + \sum_{k=2}^{K} \sum_{k'=2}^{K} \pi_k \pi_{k'} D_k D_{k'}\phi(0, t_2,..., t_K) \qquad (10)$$

where the time subscript t should be distinguished from the real arguments of $\phi(\cdot)$, $D_k\,\phi(\cdot) = \partial\phi(\cdot)/\partial t_k$, $D_k^2\,\phi(\cdot) = \partial^2\phi(\cdot)/\partial t_k^2$ and $D_k D_{k'}\phi(\cdot) = \partial^2\phi(\cdot)/\partial t_k\,\partial t_{k'}$.

If $(Y_t, X_{2t},..., X_{Kt})$ is a multivariate normal, it is well-known that the conditional expectation and conditional variance of any of these variables, given the remaining variables, are respectively linear in and independent of the conditioning vector (see (*48*, p. 523)). Although sufficient for the existence of these conditional expectations and variances, multivariate normality is by no means necessary, as KLR's lemma shows.

KLR's lemma provides conditions for the existence of a linear reduced-form equation (or a linear population regression function) between an endogenous variable, Y, and a set of exogenous variables, $X_1,..., X_K$. In light of KLR's lemma, Granger's definitions of causality and Sargent's definition of exogeneity are clearly inadequate. The inequalities between predictive variances stated in Granger's definitions and the lack of correlation between the innovation (of a covariance-stationary, purely indeterministic and invertible process) and another variable (which follows a covariance-stationary, purely indeterministic and invertible process) stated in Sargent's definition are not sufficient for the existence of conditional expectations or linear population regression functions among the economic variables.

The foregoing discussion provides the background for criticizing an econometric practice. Goldberger (*29*, pp. 380-88) reviews the reduced-form, recursive-form, and structural-form approaches to specify the population regression equations of endogenous variables on exogenous or predetermined variables. As he indicated in 1964 (*29*, pp. 386-87), each structural equation is intended to represent some aspect of the behavior of an economic unit, such as an individual, a firm, a sector, or a market. That the structural-form approach is a natural one in economics is demonstrated repeatedly in the large body of empirical literature in which models are built up equation by equation and unit by unit (see (*65*, pp. 468-83)). If the structural model is linear, under certain conditions we can derive an explicit reduced-form model (see (*29*, pp. 297-98)). Otherwise, we can only assume—incorrectly perhaps—the existence of an appropriate reduced-form model (as in (*24*)). Without sufficient *a priori* restrictions, the structural-form parameters will not be identified in either linear or nonlinear cases.

It is vital to realize that, in the linear case, KLR's lemma points to a possible danger inherent in using

[9] Conditions for the existence of these conditional expectations are given in (*48*, pp. 96-97).

a priori restrictions on the structural parameters because they may contradict the conditions of KLR's lemma and thereby prevent the existence of (1) the population regression function between each endogenous variable and a set of exogenous variables and (2) the constant conditional variance of each endogenous variable, given the exogenous variables. Thus, because the π_k's are functions of the structural parameters (*29*, p. 298), the identifying restrictions on the structural parameters may imply that some of the π_k's are restricted so that the conditions of KLR's lemma are not true. To better understand this difficulty, let us consider a simultaneous equation model which, if linear, may be expressed in the general form:

$$Y\Gamma + XB = U \tag{11}$$

where Y is a TxL matrix of observations on L endogenous variables, Γ is a LxL matrix of coefficients, X is a TxK matrix of observations on K exogenous variables, B is a KxL matrix of coefficients, and U is a TxL matrix of disturbances. The elements of Γ and B are the structural coefficients (see (*65*, p. 440)).

Assuming that Γ is nonsingular, we can derive the reduced form as:

$$Y = X\Pi + V \tag{12}$$

where $\Pi = -B\Gamma^{-1}$ is the matrix of reduced-form coefficents and $V = U\Gamma^{-1}$ is the matrix of reduced-form disturbances. Equation (12) exists if the joint characteristic functions of each endogenous variable and all the exogenous variables satisfy the conditions of KLR's lemma. In this case, we can interpret $X\Pi$ as the conditional mean of Y, given X and the covariance matrix of V as the conditional covariance matrix of Y given X. Furthermore, the covariance matrix of V will be independent of X. The reduced-form matrix of coefficients, Π, will be identified if and only if X has full-column rank. The connection between structural and reduced-form coefficients can be written as:

$$\Pi\Gamma + B = 0$$

or:

$$WC = 0 \tag{13}$$

where $W = (\Pi, I_K)$ is the K x (K+L) matrix of rank K and $C = (\Gamma', B')'$ is the (K+L) x L matrix of structural coefficients. The ith equation of (13) may be written as:

$$W\underline{c}_i = 0 \tag{14}$$

where $\underline{c}_i$ is the ith column of C. Because this is a consistent system of equations, a general solution is:

$$\underline{c}_i = (I - W^- W)\underline{z}_i \tag{15}$$

where W^- is a generalized inverse of W and where $\underline{z}_i$ is arbitrary (see (*48*, p. 25)).

A priori restrictions may be exclusion restrictions stating that certain elements of $\underline{c}_i$ are zero because the variables to which they relate do not appear in the ith equation of the structural form (11), or they may be linear homogenous restrictions involving two or more of the elements of $\underline{c}_i$. In any case, *a priori* restrictions on the elements of Γ and B do not violate the conditions of KLR's lemma if they are consistent with the class of solutions in equation (15). The vector, $\underline{c}_i$, satisfying *a priori* identifying restrictions, should belong to the null space of W. Otherwise, *a priori* restrictions used to identify a structure may invalidate an interpretation of the right-hand side of each corresponding reduced-form equation (with the disturbance suppressed) as the conditional expectation of an endogenous variable, given the exogenous variables. Nonlinear structural models, incidentally, share this problem unless the identifying restrictions imposed on them are consistent with the following alternative sets of conditions which guarantee the existence of the nonlinear population regression functions of the form $E(Y_t | x_{1t},...,x_{Kt}) = g(x_{1t},...,x_{Kt}) = g(x_t)$ (*48*, pp. 96-99).

1. If F(y,x) is the joint distribution function of $(Y, X_1,..., X_K) = (Y,X)$, then the set function $\int_{R_1 x S} y dF(y,x)$, where R_1xS is the cylinder set in the (Y,X)-plane with base S in the X-plane and $S \epsilon \mathcal{B}_k$ (a Borel field of sets), is absolutely continuous with respect to $\int_S dF(x)$. Furthermore, $EY_t < \infty$.

or:

2. The sample space for the variable $(Y, X_1,...,X_K)$ is the (K+1) - dimensional Euclidean space.

To elaborate on these conditions, we hold that if $EY_t = \infty$, then the (sufficient) conditions of the Radon-Nikodym theorem for the existence of $g(x_t)$ are not true. However, in many economic applications, the sample space is the n-dimensional Euclidean space, in which case the conditional expectation of Y (the indicator function of a set $A \epsilon B$, a Borel field of sets), given $X_t=x_t$, denoted by $P(A|x_t) = E(Y_t|x_t)$, is defined for all A over a x_t-set of probability 1 because the union of exceptional x_t-sets over which $P(A|x_t)$ is not defined is of zero probability measure. This finding does not mean that there are no problems if $EY_t = \infty$ whenever the sample space is the n-dimensional Euclidean space because even if $g(x_t)$ exists, it may not be consistent with the marginal distribution of x_t. Roughly speaking, $F(y_t|x_t)$ and $F(x_t)$ are *consistent* if they are the conditional and marginal distributions corresponding to some joint distribution of (Y_t, X_t). This hypothesis follows from Kolmogorov's consistency theorem which is stated in (*48*, p. 108). If this consistency condition is not met, then the probability laws fail. By not specifying $F(x_t)$, econometricians typically ignore this consistency problem.

Goldberger (*29*, p. 380) points out that, by formulating a model, econometricians attempt to characterize a joint conditional probability distribution of the endogenous variables conditional on the values of the exogenous variables using available *a priori* information. In view of the preceding discussion, this task may not be feasible because econometricians' *a priori* information may prevent interpretation of each reduced-form equation as a regression equation if the information violates the conditions under which such an interpretation is valid. Thus, econometricians cannot succeed if their *a priori* information on the structural parameters is incoherent in the sense that it is inconsistent with conditions permitting the existence of the expectation of each endogenous variable, conditional on the values of the exogenous variables. This point confirms the importance of de Finetti's and Savage's coherency condition that must always be imposed on *a priori* distributions. Furthermore, a structural model is logically invalid and the attractiveness of the structural-form approach mentioned by Goldberger (*29*, pp. 386-87) is illusory if *a priori* restrictions on structural parameters do not permit the interpretation of the corresponding reduced-form equations as the population regression equations. In light of a landmark paper by Boland (*10*, p. 506), who argues that a logically valid model is necessary before one can produce "true" empirical results, one must view this conclusion as a fundamental objection to current econometric practice.

Pierce-Haugh Characterizations of Causality

Coming full circle, we return to Granger's definitions of causality, which appeared in our initial investigation of the definitions of causality. Now that we have fully discussed the direction of the implications of full independence, partial independence, mean independence, uncorrelatedness, and orthogonality, as well as KLR's conditions for the existence of a linear regression and a constant conditional variance, we rigorously appraise works by Pierce and Haugh (*41*), Sims (*58*), and Sargent (*54*) based on Granger's definitions of causality.

In their survey article, Pierce and Haugh (*41*) developed characterizations of Granger causality, using the time series approach and certain assumptions. One of these assumptions is that there exist transformations $X_t = T_x X_t^*$ and $Y_t = T_y Y_t^*$ of the observable variables X_t^* and Y_t^* so that (X_t, Y_t) is a bivariate, nonsingular, linear covariance stationary, purely indeterministic time series and so that X_t and Y_t are causally related in the same way that X_t^* and Y_t^* are.

Very often, Pierce and Haugh argue, T_x and T_y will consist of first-difference or seasonal-difference operators because this type of transformation is frequently (presumed to be) necessary and sufficient to render the observed series stationary. Because such transformations are linear and because the optimal predictions in terms of which causality was defined by Granger are now also linear, each causality event is true of (X^*, Y^*) if and only if it is true of (X, Y). Moreover, Pierce and Haugh argue that certain nonlinear transformations of individual variables, such as logarithms or those of Box and Cox (*12*), are also causality-preserving in the above sense.

Such statements, offered as assertions, have no logical proofs verifying their truth. If they are false, a study of the relationship between the transformed variables will tell us nothing about the relationship between the untransformed variables in which we are interested. Indeed, counterexamples may be constructed to show that the transformations T_x and T_y are not causality-preserving. For example, if Y_t^* is a nonstationary process with infinite mean (as would occur if Y_t^* followed a random walk), it is possible that the first difference of this series, $Y_t = Y_t^* - Y_{t-1}^*$, is stationary with a finite mean and displays causality with X_t. Yet, because Y_t^* has no finite mean, the variance of the prediction of Y_t^* may be infinite, in which case Granger's definitions of causality cannot apply. One should also remember that the Pierce-Haugh criterion assumes covariance stationarity. However, this is a condition on only the first two moments. Statistical independence, as described earlier, is concerned with the *whole* distribution. The direction of the implications between equations (5) and (6) indicates that differencing and Box-Cox transformations are not causality-preserving.

Furthermore, certain recent papers point to serious problems with the Box-Cox transformation. In their book Box and Jenkins (*13*) argue that, given $Y_t = Y_t^*$, the transformation $Y_t^{(\lambda)} = [(Y_t^{\lambda} - 1)/\lambda]$ gives a covariance stationary process for some λ and, under normality conditions, one may consider the model:

$$\phi(B)\,\Delta^{d_1}\,\Delta_s^{d_2}\left(\frac{Y_t^{\lambda} - 1}{\lambda}\right) = \theta(B)a_t,$$

$$a_t \sim N(O, \sigma_a^2) \tag{16}$$

where B is the backward shift operator, $\Delta = I - B$, $\Delta_s = I - B^s$, $d_1 > 0$, $d_2 > 0$, $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p$, $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 ... - \theta_q B^q$, and the roots of $\phi(z) = 0$ and $\theta(z) = 0$ lie outside the unit circle where z is a complex variable.

A paper by Poirier (*44*) elaborates on the Box-Cox transformation. First, equation (16) requires the condition that $Y_t > 0$. Thus, if $(Y_t + \mu) > 0$ for some $\mu > 0$, the Box-Cox transformation can always be made on $(Y_t + \mu)$. However, if μ is unknown, the maximum likelihood estimates of the parameters of equation (16) for $Y_t + \mu$ may not exist, and the effects of μ on estimating λ and orders p, q, d_1, and d_2 become unknown. The question then arises how to assess the causality relationship among the original variables in equation (16) when μ is unknown.

On a related matter, Poirier and Melino (*45*) have shown that $E(Y_t^{(\lambda)}) = \infty$ if $-1 \leqslant \lambda < 0$ and $\mathrm{Var}(Y_t^{(\lambda)}) = \infty$ if $-2 \leqslant \lambda < 0$ for the normal Y_t. Their conclusion is important because the concept of Granger causality is not appropriate if $E(Y_t^{(\lambda)}) = \infty$.

When $\lambda \neq 0$, the density for Y_t corresponding to the normal density for $Y_t^{(\lambda)}$ will usually be that of a truncated normal and Box and Cox's likelihood function will be incorrect. Recognizing this problem, Amemiya and Powell (*1*) assumed that the untransformed variable followed a two-parameter gamma distribution and then studied the limiting behavior of the Box-Cox (incorrect) maximum likelihood estimator both for the identically and independently distributed (i.i.d.) case and the regression case. Although they acknowledge that their results were based on the assumption of the gamma distribution and thus might not be universally true, "they do point to the possible danger of using the Box-Cox method." Altogether, the weight of these various studies analyzing the properties of the Box-Cox transformation cast considerable doubt on its ability to transform two time series without distorting a causal relationship between them.[10]

In another section of their paper, Pierce and Haugh (*41*) developed a test for instantaneous causality. They argued that one can determine instantaneous causality by individually prewhitening the two series of interest, using linear one-sided filters and then by analyzing the contemporaneous cross-correlation of the two created innovation series. However, Price (*46*) has constructed two counterexamples to show that the existence of instantaneous causality is neither necessary nor sufficient for a nonzero contemporaneous cross-correlation. As Price (*46*, p. 256) states, "[t]his implies that a number of the [proofs] . . . presented by Pierce and Haugh . . . concerning the relationship between

---

[10] See (7) for a further discussion and other limitations.

the causal patterns of two time series and the restrictions on the cross-correlations of the corresponding 'whitened' series are either invalid or in need of further justification." Replying, Pierce and Haugh (*42*) conceded their earlier mistake, but maintained that the contemporaneous cross-correlation coefficient is a useful indicator of instantaneous causality when feedback from X to Y is not present. Their argument is unclear to us as no proof is given. Furthermore, in a recent paper, Evans and Wells (*20*) amend the set of equivalent and sufficient conditions under which Y does not cause X, provided by Pierce and Haugh (*41*).

In answer to Pierce and Haugh's statement that a nonlinear transformation such as autoregressive integrated moving average (ARIMA) modeling preserves causality relationships, an important paper by Schwert (*55*) uses three counterexamples to demonstrate that causal relationships among the innovations can be quite different in pattern and magnitude from the relationships among the original variables, depending upon the ARIMA models chosen to represent the variables. By implication, the Box-Jenkins methods are also not causality-preserving.

As Schwert (*55*, p. 81) points out, the use of estimates of the residuals from ARIMA models, necessitated by lack of observations on the true innovations, is analogous to an errors-in-variables approach which leads to another problem:

> If the original variables, Y_t and X_t, are measured with error, the measurement errors will generally have a different influence on the estimators of the relationship between the innovations than on the estimators of the relationship between the original variables. . . . Thus, if the original variables are measured with random errors, causality tests based on the estimated innovations series could fail to detect relationships that would be detected using the untransformed data.

There is certainly no pat procedure for choosing the proper specification of an ARIMA model. Box and Jenkins' method is, as honest practitioners readily acknowledge, "an art form." Pindyck and Rubinfeld (*43*, p. 473) state that "it is important to realize that the specification of an ARIMA model is an art, rather than a science," while Granger and Newbold (*28*, p. 107) affirm that "it remains the case that there does not exist a clearly defined procedure leading in any given situation to a unique identification." The basic problem is that the ARIMA models are not logically valid unless specific assumptions are true (see (*61*, p. 139)). As in the case of many assumptions, the truth of assumptions underlying ARIMA models cannot be determined *a priori.*

A related problem with Box and Jenkins' methods is that the sample autocorrelation function will not accurately reflect the properties of the population autocorrelation function (see (*47*, p. 331)). As a result, a researcher could easily misidentify some model as an ARIMA process.

One should stress that, however elaborate one's assumptions (or wishes), it is impossible to ascertain whether the time series sample (or some transform thereof) is from a covariance-stationary process because samples are finite and covariance-stationary processes are infinite in length. Thus, one may choose a sample that appears to be covariance-stationary, whereas a larger sample would show this not to be the case. In this regard, Tukey (*66*, p. 50) has proved that any "finite-extent function can arise, to an arbitrarily close approximation, as a sample from a process with any spectrum." One cannot distinguish among infinite-duration processes on the basis of a finite-length time series without making strong assumptions whose truth we do not know.

Finally, there is a logical problem with Box and Jenkins' method of determining the order q of the moving-average part of an ARIMA model. The moving-average process of finite order q has an autocorrelation function which is zero beyond the order q. It is incorrect to conclude from this that, given the jth autocorrelation coefficient, $\rho_j \neq 0$ for j = 1, 2,..., q and $\rho_j = 0$ for $j > q$, the process has a moving-average representation. The condition that a real valued series (Y_t) has a nonzero autocorrelation of order q and no nonzero autocorrelation of order greater than q is necessary, but *not* sufficient, for Y_t to have a moving average representation (see (*51*, lemma 1)). If

one looks at a sample autocorrelation function, which happens to have a cutoff after lag q and concludes that a moving average model of order q is appropriate for the series, then one would be erroneously treating a necessary condition as if it were a sufficient condition.

Sims-Granger Causality Testing

Sims (*58*) proved two theorems (also described in Sargent's book (*54*)) that provide the basis for his causality test. Theorem 1 states: Let (X_t, Y_t) be a jointly-covariance-stationary-strictly-indeterministic-process with mean zero. Then (Y_t) fails to Granger cause (X_t) if and only if there exists a vector-moving-average-representation of the form:

$$\begin{bmatrix} X_t \\ \\ Y_t \end{bmatrix} = \begin{bmatrix} C_{11}(B) & 0 \\ \\ C_{21}(B) & C_{22}(B) \end{bmatrix} \begin{bmatrix} \epsilon_t \\ \\ U_t \end{bmatrix} \qquad (17)$$

where ϵ_t and U_t are serially uncorrelated processes with means zero and $E\epsilon_t U_s = 0$ for all t and s. In addition, the one-step ahead prediction errors:

$$X_t - E(X_t | X_{t-1}, ..., Y_{t-1}, ...)$$

and:

$$Y_t - E(Y_t | Y_{t-1}, ..., X_{t-1}, ...) \qquad (18)$$

are each linear combinations of ϵ_t and U_t.

Theorem 2 of Sims states: Y_t can be expressed as a distributed lag of current and past X's (with no future X's) with a disturbance process that is orthogonal to past, present, and future X's if and only if Y does not Granger cause X. That is:

$$Y_t = \sum_{j=0}^{\infty} b_j X_{t-j} + A_t \qquad (19)$$

where $E(A_t X_s) = 0 \ \forall(t,s)$ if and only if Y does not Granger cause X.[11] Sims uses these theorems to develop a test of Granger causality. His method is to regress Y_t on all X's:

$$Y_t = (...X_{t+1}, X_t, X_{t-1}, ...) + V_t \qquad (20)$$

A researcher then tests the joint hypothesis that coefficients of all future X's are zero.

Our first comment on this test is that equation (17) is an infinite order process. In practice, one can only estimate a model of the form (20) with a finite number of independent variables. Unfortunately, truncation of lag and lead lengths of model (20) destroys the logical validity of the model in the sense described by Boland (*10*). Indeed, in view of Boland's (*11*, p. 85) demonstration that there is no valid *approximate modus ponens*, the conclusions given by a truncated model of the form (20) cannot be approximately true, even when the truncated model is approximately true.

Second, the procedure proposed by Sims is a test of only a necessary, but not a sufficient, condition for Granger noncausality. The reason is that the lower triangularity restriction on the coefficient matrix of equation (17) only implies the condition that the coefficients of the future values of X in equation (19) are zero. The restriction does not imply the condition that $E\epsilon_t U_s = 0$ for all t and s or $E(A_t X_s) = 0 \ \forall(t,s)$ (see (*52*)). Even if we reject a necessary condition for Granger noncausality on the basis of Sims' test, the probability that Granger noncausality is false is less than 1 because conclusions of statistical tests do not hold with probability 1. A statement claiming that Granger's causality holds with probability less than 1 is thus neither absolutely true nor absolutely false!

In large samples, the situation is even worse because the power of Sims' test does not go to 1 as the sample size goes to infinity (see (*52*, p. 407)). Behind Sargent's conclusion that Sims' test may fail to reject the hypothesis in infinite samples, even when it is false, is an identification problem corresponding to an infinite duration process (see (*61*, pp. 140-41)). Gabrielsen (*23*) presented an important proof that the existence of a consistent estimator $\hat{\theta}$ for a parameter θ is a sufficient condition for its identifiability. An equivalent statement is that identifiability is a necessary condition for consistency. If a parameter is not identifiable in a model, then it has no consistent estimator, and consistent tests of hypotheses about

[11] Recall that the condition $E(A_t X_s) = 0 \ \forall(t,s)$ does not imply that $E(A_t | X_s) = EA_t = 0$ which is required to show that X_t is econometrically exogenous with respect to Y_t (see the discussion after equation (7)).

the parameter do not exist. Therefore, without additional restrictions on the coefficients and the covariance between ϵ_t and U_t, the model (20) is not identified. Tukey (*66*, p. 50) adds that:

> the existence of such a difficult connection between observables and infinite-duration processes is, for me, a good reason to doubt the adequacy of a logical structure focused on infinite duration processes to guide the analysis of data .... We cannot know precisely what the spectrum is if we know only the finite-length process, even exactly. Our fate in the real world is worse, of course, since we cannot know even the finite-length process exactly.[12]

For further discussion on spectral estimation, see (*3*).

General Remarks on Causality Testing

A common problem with any of the causality tests described is that the simple bivariate models can obscure more subtle (and not so subtle) relationships involving other variables. When two events are the effects of a third event which is the cause of them, logicians describe the causal relationship between the two events as the "*fallacy of the common cause.*" This is a problem acknowledged by proponents such as Granger (*26*), Pierce (*40*), and Sims (*56, 57*) and is analyzed by Jacobs, Leamer, and Ward (*35*) who show that "any specification error renders the causality tests uninterpretable." Not only can causality tests reject exogeneity when the variable is exogenous because of the identification problems mentioned above, it can also accept exogeneity when the variable is, in fact, endogenous.

The stationarity assumption used by Sims (*58*) and Sargent (*53*) is inappropriate for aggregate time series. This problem can be seen from Swamy, Barth, and Tinsley (*61*, pp. 133-36) who prove that aggregation over disparate micro relations can yield models with time-varying coefficients, a result that is not always appreciated in either time series or conventional econometric literature. As shown by Swamy and Tinsley (*63*), a time-varying parameter model can accommodate a great variety of nonstationary processes. Also related to this argument is the Lucas critique; namely, when structural parameters are not invariant under alternative policy regimes, the stationarity assumptions used by Sims and Sargent are not reasonable.

Some Thoughts on Causality and Related Topics

In a wide-ranging, yet cogent, essay on the nature of causation, Zellner (*70*) argues articulately about the inadequacy of Granger's definition of causality and the superiority of the philosophical definition of causality provided by Feigl for econometric work. According to Feigl, the concept of causation is defined in terms of *predictability according to a law* (or more properly, according to a set of laws) (see (*70*, p. 12)). The reason Zellner (*70*, p. 51) prefers Feigl's definition of causation to all the other definitions he considers is that departures from Feigl's definition have produced problems, while offering little in the way of dependable and convincing results. Zellner (*70*, p. 51) further points out that in establishing and using economic laws in econometrics one can have little doubt that economic theory, data, and other subject matter considerations, as well as econometric techniques including modern time series analysis, must all play a role.

Although we agree with Zellner's views, Blaug's statement (*9*, pp. 160-62) concerning economic laws also deserves some attention. In Blaug's view, the term "law" has gradually acquired an old-fashioned ring and economists now prefer to present their most cherished general statements as "theorems" rather than as "laws." He further says:

> At any rate, if by laws we mean well-corroborated, universal relations between events or classes of events deduced from independently tested initial conditions, few modern economists would claim that economics has so far produced more than one or two laws.

[12] Other papers by Jacobs, Leamer, and Ward (*35*), Engle, Hendry and Richard (*19*), and Buiter (*14*) have discussed this subject and suggested that there is a problem of testing for exogeneity. However, none has discussed the identification problem with any degree of comprehensiveness.

The statement is accompanied by the following illuminating footnote:

> Samuelson . . . remarks that years of experience have taught him how treacherous are economic "laws" in economic life: e.g. Bowley's Law of constant relative wage share; Long's Law of constant population participation in the labor force; Pareto's Law of unchangeable inequality of incomes; Denison's Law of constant private saving ratio; Colin Clark's Law of a 25 percent ceiling on government expenditure and taxation; Modigliani's Law of constant wealth-income ratio; Marx's Law of the falling rate of real wage and/or the falling rate of profit; Everybody's Law of a constant capital-output ratio. If these be Laws Mother Nature is a criminal by nature.

As indicated earlier, some econometric assumptions have become so dear that they have assumed a power nearly as compelling as law. Thus, if stationarity for the transformation of the variable Y_t in equation (16) (given some d_1, d_2 and λ), is taken to be a law, then Mother Nature must surely be a scofflaw.

In view of these statements, a more modest, but more realistic, approach might be to define causation in terms of "predictability according to a sufficient and logically consistent explanation or theory."[13] The qualification "sufficient and logically consistent" is added to indicate that, at the very minimum, real economic theories must be logically valid if they are to provide "true" explanations of real economic phenomena. This requirement holds even though the logical validity of any explanation does not imply its truth. Nevertheless, consistency of knowledge plays a major role in how one explains the world; the *truth* of knowledge is much more difficult to ascertain (see (*10*)). A modest research program then becomes: if all the predictions of a logically valid theory pass a conventional test (of observation), then we may say without contradiction that the theory is so far confirmed.

Swamy, Barth, and Tinsley (*61*, pp. 131-36) make serious efforts to exploit economic theories in empirical research by using a minimal set of auxiliary assumptions and coherent prior information. In their expectations model, offered as an alternative to rational expectations, subjective probabilities are not carelessly equated to "objective probabilities" and all regression coefficients are allowed to vary over time as a concession to Samuelson's ironic list of so-called laws. We sometimes prefer the above model because (1) it avoids Box and Jenkins', Pierce and Haugh's, and Sims and Sargent's stationarity assumptions or stationarity-inducing transformations and (2) it is not forced to rely on econometric assumptions about *a priori* structural parameter information that may contradict necessary and sufficient conditions for the existence of the conditional expectations of endogenous variables, given the exogenous variables. Furthermore, deviating from usual practice, the model does not confine all uncertainty to the intercept term, but allocates it over all coefficients in each equation. Because the model is less restrictive, this procedure of first distributing uncertainty to all coefficients and then of letting data determine the major channels of uncertainty is less objectionable than the conventional procedure which first arbitrarily allocates all uncertainty to the intercept term and then forces the data to satisfy this restriction (see (*49*) for a survey of initial efforts in this research program and also (*50*, *63*, *64*) for some of the latest theoretical and empirical results).

In the above model, the conditions for logical validity are weaker than those which derive ARIMA and conventional econometric models. Because the problem of induction is unsolved, logical validity requires that the truth of one's premises or assumptions must be assumed.[14] Under these circumstances, it is prudent to work with a minimal set of assumptions. How compelling the above advice is depends, of course, on the purpose of a model. If forecasting future events is the single object of a modeling endeavor, then predictive success is a sufficient argument in favor of the model. This view of the *role* of

[13] Perhaps by "law" Zellner (*70*) meant a "sufficient and logically consistent explanation or theory."

[14] For a demonstration that causality proponents have fallen into the trap of attempting to solve the well-known "problem of induction," see (*62*).

models is called "instrumentalism" (see *10*, p. 508)). In this case, *a priori* truth of the assumptions is not required *if* it is already known that the predictions are true or acceptable by some conventional criterion (see (*10*, p. 509)). In contrast, those economists who see the object of science as finding the *one* true theory of the economy will find their task difficult, if not impossible. On the surface, instrumentalism offers a valid guide for scientific investigation. It is unfortunate that no single model predicts all variables better than all other models for all time periods. This predictive criterion must eventually exhaust itself. Because it is impossible to foretell the time of failure, we cannot even pick a model based on instrumentalism. However, we can reject models on the grounds of logical invalidity, as we did in the preceding sections.

Given the difficulty of choosing among logically valid models, the principle of parsimony has sometimes been invoked as a tempting guide. The imposition of certain restrictions on the time-varying parameter models can lead to conventional regression models with heteroscedastic or serially correlated error terms (or the ARIMA models) (see (*63*, pp. 107-08)). Although these restrictions produce substantial economies in parameterizing a model, such economies are not without cost. Despite its tempting name, the principle of parsimony—preferring restricted specifications to more complex modeling whenever the performance of the former in prediction is *almost as good* as that of the latter—has little justification unless the conventional or ARIMA models perform *at least as well* as some more general model, for example, the alternative expectations model proposed by Swamy, Barth, and Tinsley (*61*).

The conventional models, including ARIMA models, exhibit episodic breakdowns and perform poorly in prediction. The usual practice is to repair such models by extensive respecification or, more often in the shorter run, with judgmental "add factors," dummy variables, and "ratchet" arguments. Following Lakatos (see (*9*, p. 36)), we may call this research practice "degenerating" because it involves endlessly adding *ad hoc* adjustments that merely accommodate whatever new facts become available. A positive contribution is possible only if the scientific research program is *theoretically progressive*—that is, if a successive formulation of the program contains "excess empirical content" over its predecessor, that is, the program predicts "some novel, hitherto unexpected fact" or if the program is *empirically progressive*—that is, if "this excess empirical content is corroborated." The limited evidence presented by Havenner and Swamy (*32*), Resler, Barth, Swamy, and Davis (*50*), and Swamy, Tinsley, and Moore (*64*) appears to favor the claim that the time-varying coefficient models facilitate progressive scientific research programs. Just as the philosophy of instrumentalism does not permit us to call one of the existing models the best predictor of all variables for all time periods, so the principle of parsimony does not permit us to call one model the best.

Time-varying coefficient models such as those Swamy and Tinsley (*63*) propose may be too complex to be useful. Indeed, Popper has argued that theoretical simplicity may be equated to the degree to which a theory can be falsified, in the sense that the simpler the theory, the stricter its observable implications and, hence, the greater its testability. It is because simpler theories have these properties that we aim for simplicity in science. But this principle is not universally agreed upon. Thus, Blaug (*9*, p. 25) casts his doubts about Popper's notion of simplicity as follows:

> It is doubtful that this is a convincing argument, since the very notion of simplicity of a theory is itself highly conditioned by the historical perspective of scientists. More than one historian of science has noted that the elegant simplicity of Newton's theory of gravitation, which so impressed nineteenth-century thinkers, did not particularly strike seventeenth-century contemporaries, and if modern quantum mechanics and relatively theory are true, it must be conceded that they are not very simple theories. Attempts to define precisely what is meant by a simpler theory have so far failed . . . , and Oscar Wilde may have been right when he quipped that the truth is rarely pure and never simple.

One of these statements is accompanied by the following footnote:

> As Polanyi . . . has observed, "great theories are rarely simple in the ordinary sense of the term. Both quantum mechanics and relativity theory are very difficult to understand; it takes only a few minutes to memorize the facts accounted for by relativity, but years of study may not suffice to master the theory and to see these facts in its context."

Conclusions

The term, "causality," as used by Granger and his followers, has been erroneously identified with feedback or dependence and loosely with correlation (see (*71*, p. 313)). We have contrasted this new usage with traditional approaches proposed by scientific philosophers and surveyed by Zellner (*70*). By every acceptable norm, the latter approach may still offer sharper views on the definition of causation. There is evidence (see (*71*, p. 313)) that Granger himself has altered his views since his initial article. Granger now argues: "Provided I define what I personally mean by causation, I can use the term" (*27*, pp. 333 and 337). What Granger means by causality is that knowledge of Y_t increases one's ability to forecast X_{t+1} in a least squares sense. Truth, like beauty, may be in the eyes of the beholder, but it is still fair to insist that the purpose of language is to communicate and clarify. Perhaps much of the confusion surrounding the interpretation of causality tests would not have arisen if such tests had instead been labeled "tests of relative predictive efficiencies" or some other neutral terms suggested by Schwert (*55*, p.82).

More important, the difficulty with using Granger's causality definitions, even as a measure of relative forecasting efficiency, is that the same relationship may not continue into the forecast period. There is indeed every reason to believe that such a relationship will change. One may support this belief by contemplating the numerous structural upheavals of the seventies as well as the implication of Lucas' critique suggesting that individual behavior (and hence structural coefficients) will change when policy rules change.

Zellner (*70*) recommends using Feigl's definition of causation, which we respectfully modify to read," predictability according to a sufficient and logically consistent theory." This modification is necessary because contemporary economists prefer to present their most cherished general statements as theorems rather than as laws.

Causality tests were created with the best of intentions, but one must be careful never to ask more of the data than they can deliver. It is unfortunate that these tests seem to ask for more. However, if one can find a way to avoid the contradictions between the *a priori* restrictions on the structural parameters and the conditions of KLR's lemma and if these restrictions are overidentifying, then one can invoke Wu's procedures (*69*) to examine the significance of the covariances between independent variables and the disturbances (provided we have an identifiable maintained hypothesis).[15] Unlike causality tests, Wu's procedures adhere to a law of large numbers; the powers of his tests, therefore, equal 1 in sufficiently large samples.

Where, then, is the econometrician left in devising a modeling strategy to determine causality? Zellner's fundamental argument is that the soundness of our conclusion about causality is ultimately based on the soundness of economic theory to determine causality. In our view, this advice is wise, and in the spirit of Zellner's theme, we end with a revealing conversation between Fisher and Cochran, which Zellner quotes (*72*, p. 13):

> About 20 years ago, when asked in a meeting what can be done in observational studies to clarify the step from association to causation, Sir Ronald Fisher replied: "Make your theories elaborate." The reply puzzled me at first, since by Occam's razor the advice usually given is to make theories as simple as is consistent with known data. What Sir Ronald meant, as the subsequent discussion showed, was that when constructing a causal hypothesis one should envisage as many *different* consequences of its truth as possible, and plan observational studies to discover whether each of these consequences is found to hold.

---

[15] Our earlier discussion indicates that in the normal case uncorrelatedness is equivalent to mean independence.

References

(1) Amemiya, T., and J. L. Powell. "A Comparison of the Box-Cox Maximum Likelihood Estimator and the Non-Linear Two-Stage Least Squares Estimator," *Journal of Econometrics*, Vol. 17, Dec. 1981, pp. 351-81.

(2) Anderson, T. W. *The Statistical Analysis of Time Series.* New York: John Wiley & Sons, 1971.

(3) Barth, J. R., and J. T. Bennett. "Cyclical Behavior Seasonality, and Trend in Economic Time Series," *Nebraska Journal of Economics and Business*, Vol. 13, No. 1, Winter 1974, pp. 48-69.

(4) Belongia, M., and D. A. Dickey. "Prefiltering and Causality Tests," *Agricultural Economics Research*, Vol. 34, No. 4, Oct. 1982, pp. 10-14.

(5) Bessler, D. A., and J. A. Brandt. "Causality Tests in Livestock Markets," *American Journal of Agricultural Economics*, Vol. 64, No. 1, Feb. 1982, pp. 140-44.

(6) Bessler, D. A., and L. F. Schrader. "Measuring Leads and Lags Among Prices: Turkey Products," *Agricultural Economics Research*, Vol. 32, No. 1, Jan. 1980, pp. 1-7.

(7) Bickell, P. J., and K. A. Doksum. "An Analysis of Transformation Revisited," *Journal of the American Statistical Association*, Vol. 374, No. 3, June 1981, pp. 296-311.

(8) Bishop, R. V. "The Construction and Use of Causality Tests," *Agricultural Economics Research*, Vol. 31, No. 4, Oct. 1979, pp. 1-6.

(9) Blaug, M. *The Methodology of Economics.* New York: Cambridge Univ. Press, 1980.

(10) Boland, L. A. "A Critique of Friedman's Critics," *Journal of Economic Literature*, Vol. 17, No. 2, June 1979, pp. 503-22.

(11) Boland, L. A. "Satisficing in Methodology: A Reply," *Journal of Economic Literature*, Vol. 19, No. 1, Mar. 1981, pp. 84-86.

(12) Box, G. E. P., and D. R. Cox. "An Analysis of Transformations," *Journal of the Royal Statistical Society*, Series B, Vol. 26, 1964, pp. 211-52.

(13) Box, G. E. P., and G. M. Jenkins. *Time-Series Analysis: Forecasting and Control.* 2nd ed. San Francisco: Holden-Day, 1976.

(14) Buiter, W. "Granger-Causality and Stabilization Policy." NBER Technical Working Paper No. 10. 1981.

(15) Chamberlain, G. "The General Equivalence of Granger and Sims Causality," *Econometrica*, Vol. 50, No. 3, May 1982, pp. 569-82.

(16) Chung, K. L. *A Course in Probability Theory.* 2nd ed. New York: Academic Press, 1974.

(17) Conway, R. K., and J. R. Barth. "New Developments in Macroeconomic Theory: A Prospectus and Appraisal," *Agricultural Economics Research*, Vol. 35, No. 3, July 1982, pp. 23-29.

(18) Conway, R. K., and E. G. Fryar, Jr. "Rational Expectations in Agricultural Economics Research and Policy Analysis: Some Pitfalls." Univ. of Arkansas Agricultural Economics Staff Report. 1983.

(19) Engle, R. F., D. F. Hendry, and J. F. Richard. "Exogeneity," *Econometrica*, Vol. 51, No. 2, Mar. 1983, pp. 277-304.

(20) Evans, L., and G. Wells. "Pierce and Haugh on Characterizations of Causality: A Re-examination," *Journal of Econometrics*, Vol. 23, No. 3, Dec. 1983, pp. 331-36.

(21) Feige, E., and D. K. Pearce. "The Casual Causal Relationship Between Money and Income: Some Caveats for Time Series Analysis," *Review of Economics and Statistics*, Vol. 61, No. 4, Nov. 1979, pp. 521-33.

(22) Florens, J. P., and M. Mouchart. "A Note on Noncausality," *Econometrica*, Vol. 50, No. 3, May 1982, pp. 583-92.

(23) Gabrielsen, A. "Consistency and Identifiability," *Journal of Econometrics*, Vol. 8, No. 2, Oct. 1978, pp. 261-64.

(24) Gallant, A. R., and A. Holly. "Statistical Inference in An Implicit, Non-Linear, Simultaneous Equation Model in the Context of Maximum Likelihood Estimation," *Econometrica*, Vol. 48, No. 3, Apr. 1980, pp. 697-720.

(25) Geweke, J., R. Meese, and W. Dent. "Comparing Alternative Tests of Causality in Temporal Systems: Analytic Results and Experimental Evidence," *Journal of Econometrics*, Vol. 21, No. 3, Feb. 1983, pp. 161-94.

(26) Granger, C. W. J. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica*, Vol. 37, No. 3, July 1969, pp. 424-38.

(27) __________. "Testing for Causality: A Personal Viewpoint," *Journal of Economic Dynamics and Control*, Vol. 2, No. 4, Nov. 1980, pp. 329-52.

(28) __________, and P. Newbold. *Forecasting Economic Time Series.* New York: Academic Press, 1977.

(29) Goldberger, A. S. *Econometric Theory.* New York: John Wiley & Sons, 1964.

(30) __________. "Structural Equation Methods in the Social Sciences," *Econometrica*, Vol. 40, No. 6, Nov. 1972, pp. 979-1002.

(31) __________. *Topics in Regression Analysis.* New York: The Macmillan Company, 1968.

(32) Havenner, A., and P. A. V. B. Swamy. "A Random Coefficient Approach to Seasonal Adjustment of Economic Time Series," *Journal of Econometrics*, Vol. 15, No. 2, Feb. 1981, pp. 177-209.

(33) Hayashi, F., and C. A. Sims. "Nearly Efficient Estimation of Time Series Models with Predetermined, But Not Exogenous, Instruments," *Econometrica*, Vol. 51, No. 3, May 1983, pp. 783-98.

(34) Heien, D. M. "Markup Pricing in a Dynamic Model of the Food Industry," *American Journal of Agricultural Economics*, Vol. 62, No. 1, Feb. 1980, pp. 10-18.

(35) Jacobs, R. L., E. E. Leamer, and M. P. Ward "Difficulties with Testing for Causation," *Economic Inquiry*, Vol. 17, July 1979, pp. 401-13.

(36) Kagan, A. M., Y. V. Linnik, and C. R. Rao. *Characterization Problems in Mathematical Statistics.* New York: John Wiley & Sons, 1973.

(37) Ling, R. F. "Review of Correlation and Causation by David A. Kenny," *Journal of the American Statistical Association*, Vol. 77, No. 378, June 1982, pp. 488-91.

(38) Nelson, C. R., and G. W. Schwert. "Tests for Predictive Relationships Between Time Series Variables: A Monte Carlo Investigation," *Journal of the American Statistical Association*, Vol. 377, No. 1, Mar. 1982, pp. 11-18.

(39) Phipps, T. "Farmland Prices and the Return to Land: An Application of Causality Testing." Paper presented at American Agricultural Economics Association Meetings, Logan, Utah, 1982.

(40) Pierce, D. A. "Relationships—and the lack thereof—Between Economic Time Series, with Special Reference to Money and Interest Rates," *Journal of the American Statistical Association*, Vol. 72, No. 357, Mar. 1972, pp. 11-21.

(41) __________, and L. D. Haugh. "Causality in Temporal Systems: Characterizations and a Survey," *Journal of Econometrics*, Vol. 5, No. 3, May 1977, pp. 265-95.

(42) __________ . "The Characterization of Instantaneous Causality: A Comment," *Journal of Econometrics*, Vol. 10, No. 2, June 1979, pp. 257-59.

(43) Pindyck, R. S., and D. L. Rubinfeld. *Econometric Models and Economic Forecasts.* New York: McGraw-Hill Book Company, 1976.

(44) Poirier, D. J. "Experience with Using the Box-Cox Transformation when Forecasting Economic Time Series—A Comment," *Journal of Econometrics*, Vol. 14, No. 2, Oct. 1980, pp. 277-80.

(45) __________ , and A. Melino. "A Note on the Interpretation of Regression Coefficients Within a Class of Truncated Distributions," *Econometrica*, Vol. 46, No. 5, Sept. 1978, pp. 1207-09.

(46) Price, J. M. "The Characterization of Instantaneous Causality: A Correction," *Journal of Econometrics*, Vol. 10, No. 2, June 1979, pp. 253-6.

(47) Priestley, M. B. *Spectral Analysis and Time Series*. New York: Academic Press, 1983.

(48) Rao, C. R. *Linear Statistical Inference and Its Applications*. 2nd ed. New York: John Wiley & Sons, 1973.

(49) Rausser, G. C., Y. Mundlak, and S. R. Johnson. "Structural Change, Parameter Variation and Forecasting," *New Directions in Econometric Modeling and Forecasting in U.S. Agriculture* (ed. G. C. Rausser). New York: American Elsevier Publishing Co., 1983.

(50) Resler, D. H., J. R. Barth, P. A. V. B. Swamy, and W. D. Davis. "Detecting and Estimating Changing Economic Relationships: The Case of Discount Window Borrowings." Special Studies Paper 165. Washington, D.C.: Federal Reserve Board, 1982.

(51) Rose, D. E. "Forecasting Aggregates of Independent ARIMA Processes," *Journal of Econometrics*, Vol. 5, No. 3, May 1977, pp. 323-46.

(52) Sargent, T. J. "Causality, Exogeneity, and Natural Rate Models: Reply to C. R. Nelson and B. T. McCallum," *Journal of Political Economy*, Vol. 87, 1979, pp. 403-09.

(53) __________ . "A Classicial Macroeconometric Model for the United States," *Journal of Political Economy*, Vol. 84, No. 2, Apr. 1976, pp. 207-37.

(54) __________ . *Macroeconomic Theory*. New York: Academic Press, 1979.

(55) Schwert, G. W. "Tests of Causality: The Message in the Innovations," *Carnegie-Rochester Conference Series on Public Policy*, Vol. 10 (ed. Karl Brunner and Allan H. Meltzer). Supplement to *Journal of Monetary Economics.* Amsterdam: North-Holland Publishing Company, 1979, pp. 55-96.

(56) Sims, C. A. "Exogeneity and Causal Ordering in Macroeconomic Models," *New Methods in Business Cycle Research: Proceedings from a Conference* (ed. C. A. Sims). Minneapolis: Federal Reserve Bank of Minneapolis, 1977, pp. 23-43.

(57) __________ . "Macroeconomics and Reality " *Econometrica*, Vol. 48, No. 1, Jan. 1980, pp. 1-49.

(58) __________ . "Money, Income, and Causality " *American Economic Review*, Vol. 62, No. 4, Sept. 1972, p. 540-52.

(59) Slaugh, J. R. "Granger-Sims Causality: A Brief Survey of Its Use and Misuse." Unpublished manuscript. National Science Foundation, 1981.

(60) Spreen, T. H., and J. S. Shonkwiler. "Causal Relationships in the Fed Cattle Market," *Southern Journal of Agricultural Economics*, Vol. 13, No. 1, July 1981, pp. 149-53.

(61) Swamy, P. A. V. B., J. R. Barth, and P. A. Tinsley. "The Rational Expectations Approach to Economic Modelling," *Journal of Economic Dynamics and Control*, Vol. 4, No. 2, May 1982, pp. 125-47.

(62) Swamy, P. A. V. B., R. K. Conway, and P. von zur Muehlen. "The Foundations of Econometrics—Are There Any?" Special Studies Paper 182. Washington, D.C.: Federal Reserve Board, 1984.

(63) Swamy, P. A. V. B., and P. A. Tinsley. "Linear Prediction and Estimation Methods for Regression Models with Stationary Stochastic Coefficients," *Journal of Econometrics*, Vol. 12, No. 2, Feb. 1980, pp. 103-42.

(64) __________, and G. Moore. "An Autopsy of a Conventional Macroeconomic Relation: The Case of Money Demand." Special Studies Paper 167. Washington, D.C.: Federal Reserve Board, 1982.

(65) Theil, H. *Principles of Econometrics.* New York: John Wiley & Sons, 1971.

(66) Tukey, J. W. "Comments on Seasonality: Causation, Interpretation, and Implications by C. W. J. Granger," *Seasonal Analysis of Economic Time Series* (ed. A. Zellner). U.S. Department of Commerce, Bureau of the Census, 1978, pp. 50-53.

(67) Ward, R. W. "Asymmetry in Retail, Wholesale, and Shipping Point Pricing for Fresh Vegetables," *American Journal of Agricultural Economics*, Vol. 64, No. 2, May 1982, pp. 205-12.

(68) Whittle, P. *Probability.* New York: John Wiley & Sons, 1976.

(69) Wu, D. M. "Alternative Tests of Independence Between Stochastic Regressors and Disturbances," *Econometrica*, Vol. 41, No. 3, July 1973, pp. 733-50.

(70) Zellner, A. "Causality and Econometrics," *Carnegie-Rochester Conference Series on Public Policy.* Vol. 10 (ed. Karl Brunner and Allan H. Meltzer). Supplement to *Journal of Monetary Economics.* Amsterdam: North-Holland Publishing Company, 1979.

(71) __________. "Comment," *Journal of the American Statistical Association*, Vol. 77, No. 378, June 1982, pp. 313-14.

(72) __________. *An Introduction to Bayesian Inference in Econometrics*. New York: John Wiley & Sons, 1971.

Research Review

The Elasticity of Substitution and Land Use in Agricultural Production: A Cause for Optimism?

By Thomas Lutton*

Introduction

Neo-Malthusian futurists contend that world population growth is outstripping the growth of world food production and distribution. Such pessimism is reinforced by others who contend that rising resource prices, such as land, energy, and water, and resource availability constraints compounded by technological stagnation will exacerbate food scarcity in the middle to long term. Still others, while admitting to technological advances in input development and food supply, argue that the political situation in *both* developing and developed countries will occasionally result in policies precluding food distribution in times of need. Examples include output-restrictive agricultural policies in developed countries and military purchases by developing countries in times of food shortages. "Plentyists," more optimistic counterparts by contrast, contend that a series of technological advances in plant and livestock genetics, agricultural chemical breakthroughs, information dissemination through computers, and other unforeseen technological advances will mitigate the degree of the food scarcity problem. Some also contend that resource prices will decline relative to output prices, lowering the cost of producing food.

The opinions of both groups, optimists and pessimists, are well reflected in the 1981 Yearbook of Agriculture, *Will There Be Enough Food?* Both groups cite historical evidence to support their arguments, and it is difficult to reconcile their differences.

In this article, I examine a narrow portion of what appears to be one source of disagreement; that is, precisely what is meant by technology and how do we measure it? If we can agree on a definition and on the feasibility of measuring the concept, we can then ask: "Given existing technology, can domestic agriculture increase output sufficiently to provide a target output level by the year 2000?" I use a hypothetical example for heuristic purposes to illustrate the importance of input substitution when one answers this question. To the extent that input substitution is possible within existing technology, farmers' ability to cope with the price changes in selected inputs is enhanced. After an input price change, farmers' costs of production, average and marginal, are higher when their technology reflects the potential for limited input substitution. Indeed, input substitution potential is critical to understanding the problem of agricultural capacity.

I do not attempt to measure substitution potential in this article. Measurement problems are difficult given existing analytical techniques and data. I do, however, provide a general definition for technology which is identical with a production function and the underlying optimization process. Furthermore, I illustrate the importance of factor substitution in agricultural crop production by permutating an elasticity of substitution in a hypothetical constant elasticity of substitution (CES) production (cost) function (see appendix). I hope this article will sharpen the debate on agricultural capacity by calling attention to input substitution potential.

Technological Characterization

In this article, I define technology in agricultur as follows:

> Technology is a knowledge of production possibilities which individual farmers use in the purposeful application of any or all sciences (agronomy, soil science, and botany) as well as "technics" (engineering, economics, and industrial management) in the production of food and fiber.

*The author is currently a principal analyst for the Congressional Budget Office. This article was initially prepared as a contribution to the ERS world food study, 1983. At the request of the editor, the article was modified for this journal. The author wishes to thank Clark Edwards and anonymous reviewers for their helpful comments.

Economic theory suggests that a competitive producer employs this knowledge in the optimization of an objective function with a given set of relative input and output prices and technological constraints. The knowledge base may differ from farmer to farmer. Similarly, the objective function and technological constraints including environmental factors such as soil organic content, soil moisture, temperature, and pest infestation may also differ.

The parameterization of this knowledge base and constraints is contained in the functional form of the production or transformation function, which is defined as a schedule of maximum output(s) for all input combinations. Blending a simple objective function such as cost minimization with this production or transformation function, one can embody both the objective and production function into a cost function parameterization of the knowledge base. For simplicity, let total crop output be represented by Q, nonland input prices by P_N, and rental land prices by P_L. Assume that output is obtained as a function of land and nonland inputs. The minimum cost associated with producing a fixed level of Q denoted $\overline{Q}$ given a fixed set of factor prices, $\overline{P}_N$ and $\overline{P}_L$, is a scalar given by:

$$\text{MIN } C = \overline{P}_N N + \overline{P}_L L + \lambda(\overline{Q} - f(N,L)) \qquad (1)$$

where f(N,L) is the production function. N and L are input quantities of nonland and land inputs and the variables of choice used in producing any given level of output. λ is a Lagrangian multiplier. Solving equation (1) for the closed form solutions associated with N and L, we obtain a cost function, equation (2), (*8*)[1] which represents the minimum cost of producing at all output levels Q for all input prices P_N and P_L:

$$C = P_N N(P_N, P_L, Q) + P_L L(P_N, P_L, Q)$$

$$= C(P_N, P_L, Q) \qquad (2)$$

[1] Italicized numbers in parentheses refer to items in the References at the end of this article.

Substitution

I contend that the processes for future food production in North America and the resources they utilize are not immutably fixed. Minimum tillage, crop rotations, irrigation, and general input juggling within production processes can all alter input requirements for a fixed level of output. Such substitution, although often difficult to measure, minimizes adverse economic impacts of resource constraints and input price changes.[2] Although the flexibility of a single farmer after planting is limited, the set of production possibilities and alternatives before planting may be quite large. The substitution between inputs in neoclassical production theory may be viewed in numerous ways. In this analysis, I link substitution to the concept of derived demand for inputs given an output level. Inputs may be substituted for each other while the costs of producing a given output level are minimized. If the substitution potential between two inputs is zero, the average product of each in equilibrium is a constant, a result well known to input-output analysts. A casual look at the input-output measures from 1965 to 1980 demonstrates how the factor intensities have changed (see table). Note that the average products of land and labor increased 5.4 and 110.6 percent, respectively, between 1965 and 1980. The average products of agricultural chemicals and machinery decreased 46.4 and 8.6 percent, respectively, between 1965 and 1980.

Unless these variations in input-output indexes overtime are attributed solely to weather or technological change, one has difficulty explaining such changes without considering factor substitution. Factor substitution is prompted by changing relative input prices. Substitution effects must be separated from technological change, however, when such data are examined. Technological change is most evident in agricultural equipment, hybrid seed varieties, and an overall increase in the knowledge base. To separate the effects of substitution from technical change, Ray (*8*), Lopez (*6*), Huffman and Evenson (*4*), and Binswanger (*2*) find econometric

[2] The difficulty in measurement is attributed primarily to lack of homogenous input quantity data such as land and agriculture chemicals. There are also the inherent difficulties in econometrically estimating substitution potential and technology change from time series, aggregate data. For a discussion of these problems, see (*3*).

evidence of factor substitution in North American agriculture by using both time series and cross-sectional data. These studies and others (see (*10*)) find evidence of input price sensitivity conditional upon output levels, another indication of substitution potential. The greater the substitution potential the more flexibility farmers have in switching inputs and the more sensitive they are likely to be to input price changes. Simply put, there is statistical evidence that North American farmers employ different input mixes when relative prices dictate economic adjustments. In short, farmers have indeed demonstrated flexibility in their production methods.

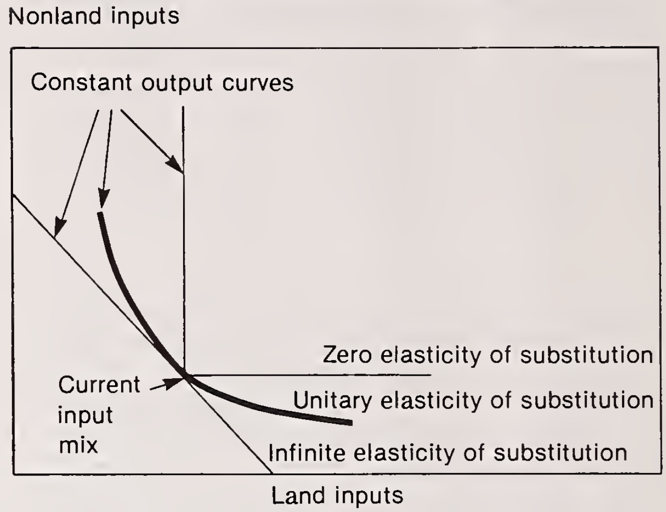
Input/Output Indexes Agricultural Production, 1965-80

Time	Acres harvested/ output	Labor hours/ output	Agriculture chemical/ output	Machinery/ output
		1965 = 1.0		
1965	1.00	1.00	1.00	1.00
1970	.95	.79	1.49	1.03
1975	.97	.59	1.46	1.03
1980	.95	.47	1.86	1.09

Source: (*11*).

Figure 1

The Elasticity of Substitution Concept[1]



[1]The elasticity is formally defined for the single output two input case as:

$$\sigma = -\frac{\partial \ln (L/N)}{\partial \ln (MP_L/MP_N)}$$

where L is the land quantity, N is nonland quantity, and MP_L and MP_N refer to the marginal products of land and nonland inputs used to produce output.

In the language of production economics, different assessments of farmer flexibility can be phrased as disagreement about the numerical value of the elasticity of substitution, *ceteris paribus.* The elasticity of substitution between two inputs is a measure of the ease or difficulty of substituting one input for another while maintaining output, given the existing technology. When inputs number more than two, the definition becomes a bit murkier. Here I confine the discussion to the measure of two inputs. However, other measures are available (see (*1,7*)).

Figure 1 illustrates the elasticity of substitution concept for two inputs. For simplicity, let there be two inputs in production of agricultural crops: land and other outputs. The other output category, hereafter referred to as nonland inputs, may be composed of labor, capital, fertilizer, seed, and so forth. For a particular moment in time identify the point, "current input mix," as one possible combination of inputs used to produce output $\bar{Q}$. The elasticity of substitution is a measure of the curvature of the constant output curves that intersect the current input mix. In our simple two-factor model, this elasticity summarizes the potential for substitution between land and other inputs. The shape of the constant product curves is affected by the number of alternative agricultural processes used to produce output. The more processes that are available, the larger the elasticity of substitution becomes; that is, the more opportunities for adjustments in input use as relative input prices change.

A Leontief production function characterized by fixed input/output equilibrium values implies a zero elasticity of substitution between land and other nonland inputs. Agricultural economists concerned with yield growth and decline would generally dispute this assumption. At the opposite extreme, where the elasticity is equal to infinity, other inputs may completely substitute for land to produce output, an implausible assumption.

Still another hypothesis is that the elasticity of substitution is unity. This hypothesis would imply that as the rental price of land rose, the value share of land (cost share) would remain at a constant share of production costs. This assumption is often embodied in Cobb-Douglas production functions. The elasticity of substitution need not equal 1, zero, or infinity. It may take on many values. For the agricultural sector the elasticity is probably nonconstant, fluctuating between 0.3 and 0.7; however, other values higher and lower may be found specific to individual crops or regions.

Assumptions Regarding Simulation

To illustrate, I choose a production function of constant elasticity of substitution, CES and its dual cost function to illustrate how the elasticity of substitution affects costs of production, crop yield, levels of prices received by farmers necessary to achieve target output and land use, given a trajectory of land prices, nonland input prices, and output levels. Let output grow at 1.32 percent per year from 1980 to 2000 so that output increases 30 percent over 1980 levels by the year 2000. Hold nonland input prices constant. To allow for land scarcity, assume the rental price for land increases at the rate of 3.5 percent per year; that is, effectively doubling between 1980 and 2000.

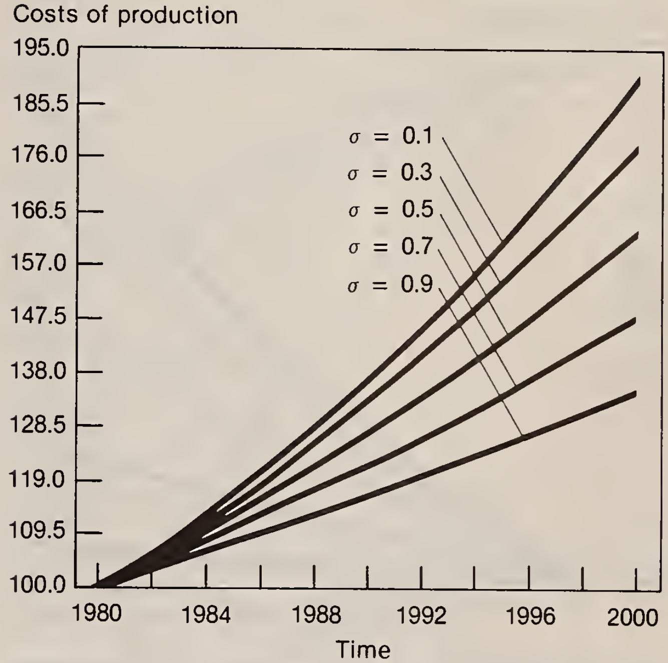
Normalizing costs, output, input quantities, yields, and average costs of production (equal to prices received by farmers in longrun competitive equilibrium) at the 1980 values equal to 100, we can simulate our simple model to illustrate the dramatic differences in magnitude of selected economic variables for the cost-minimizing farmer (see figs. 2-5). Note that figures 2-5 are internally consistent by model design. Each point on the figures corresponds to a comparative statics optimal solution. The appendix provides a more detailed discussion of the model.

Costs of Production

In figure 2, the cost of producing 30 percent more output by the year 2000 is 90.8 percent higher than 1980 levels when the elasticity of substitution (σ) is 0.1. When the $\sigma = 0.9$, costs are only 35 percent higher by the year 2000. As the elasticity grows, production costs are correspondingly lower.

Figure 2

Costs of Production with Alternative Elasticities of Substitution



For $\sigma = 0.3$, 0.5, and 0.7, the costs of production are, respectively, 78.4, 63.3, and 48.1 percent higher than 1980 levels. The lower costs are indicative of more opportunities for factor substitution between land and nonland input for the higher elasticity functions.

Prices Received by Farmers

If we assume average cost equals marginal cost and marginal cost equals price, a familiar longrun equilibrium condition, prices must rise 46.8, 37.2, 25.6, 13.9, and 3.9 percent as $\sigma = 0.1, 0.3, 0.5, 0.7, 0.9$ to entice farmers to produce 30 percent more output (fig. 3). These are substantial differences. To increase production 30 percent over 1980 levels, prices received must increase much more if the technology exhibits minimal input substitution, given the land price increase. Recall that land prices are assumed to increase by 3.5 percent per year during the 1980-2000 period, whereas nonland input prices remain at the 1980 level. The higher the elasticity of substitution, the smaller the impact of land price increases on output price. For different elasticities of substitution, this result is

Figure 3

Prices Received with Alternative Elasticities of Substitution

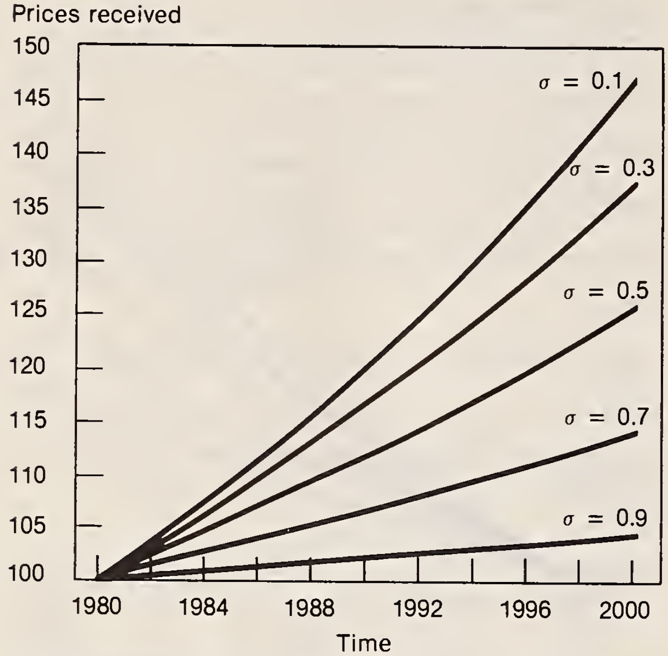
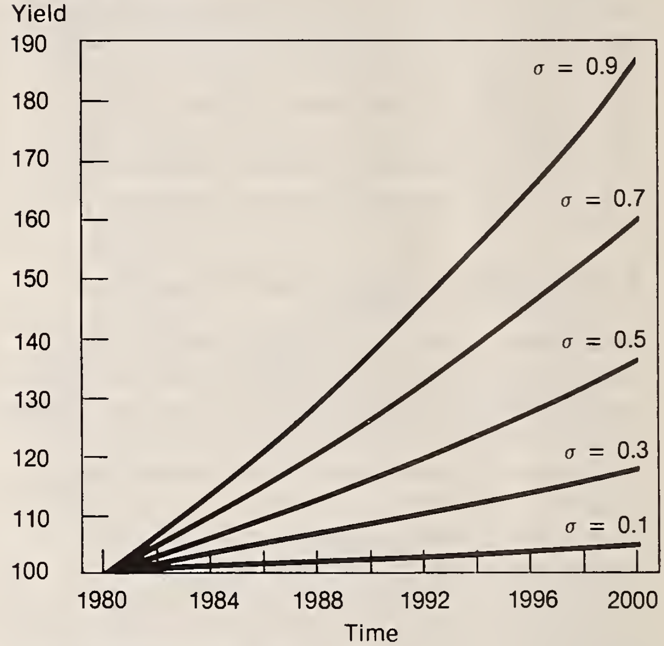


Figure 4

Yield with Alternative Elasticities of Substitution



reasonable, given the differences in costs of producing identical output levels in any given period.

Yield

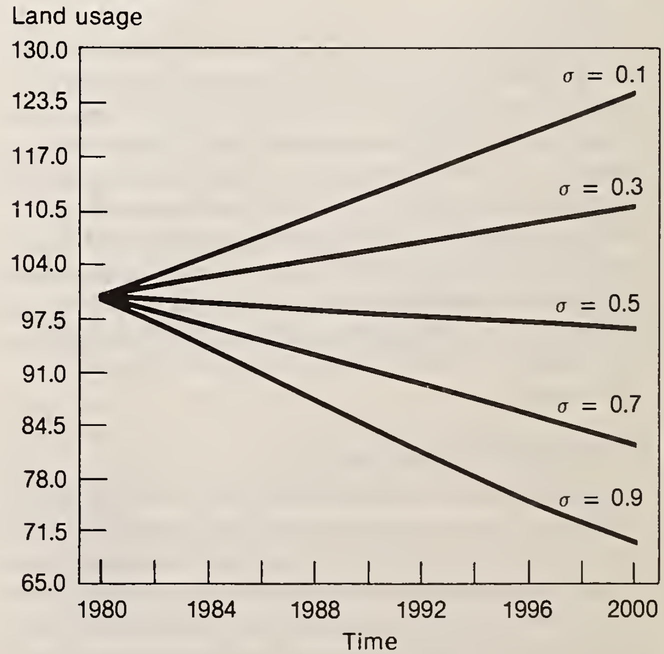
The average product of land (typically measured as yield) also exhibits marked differences for alternative values of the elasticity of substitution. When $\sigma = 0.1$, the yield for the year 2000 is only 4.7 percent higher than in 1980 (fig. 4). Alternatively, when $\sigma = 0.9$, the yield in the year 2000 is 86 percent higher than in 1980. As σ becomes larger, the average product of land increases as nonland inputs are substituted for land in producing the target level output, given the relative increase in land prices.

Land Use

In figure 5, land used to produce 30 percent more output increases 24.2 percent when $\sigma = 0.1$. Farmers must bid land away from alternatives, given the experiment preconditions. However, if $\sigma = 0.9$, land use actually declines to slightly less than 70

Figure 5

Land Usage with Alternative Elasticities of Substitution



percent of the 1980 requirement by 2000. Similarly, for $\sigma = 0.5$, land use declines over the simulation period despite the increase in output, once again illustrating the importance of the substitution measure. If for some reason it is desirable to limit land used in agricultural production for soil conservation or another reason, the higher the elasticity of substitution the fewer incentives will be required to cause farmers to switch from land to nonland inputs, an interesting implication if one is determining farmer participation in land set-aside programs.

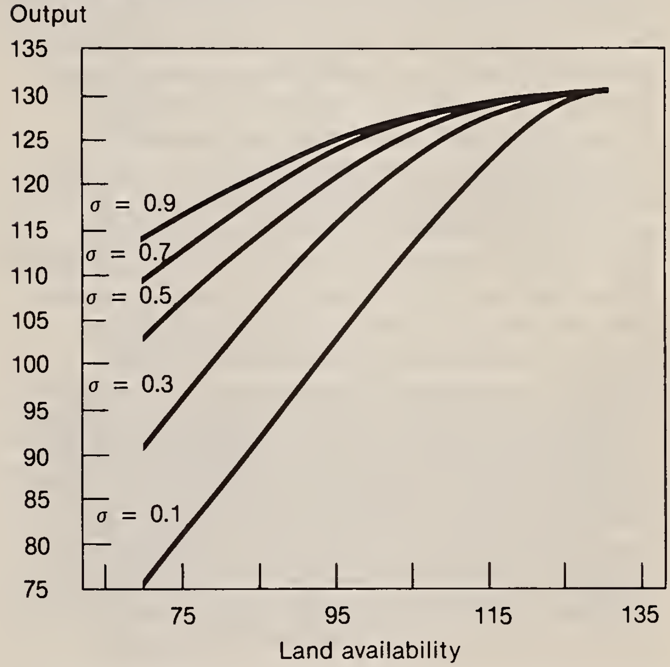
Output Effects

To illustrate the effects of a land restriction policy on output, we need to modify the model. Farm output was assumed initially to be exogenously determined. Let us relax the assumption of land price growth and hold land and nonland input prices at 1980 levels. Assume, furthermore, under constant returns to scale that a 30-percent output increase would raise land and nonland input requirements and consequently costs of production by 30 percent. Given these assumptions, consider a land policy which restricts land use, assuming a budget constraint of 130 percent of the 1980 budget. Note we are now assuming that farmers maximize output subject to a budget constraint. Farmers theoretically may substitute nonland inputs for land in an attempt to maximize output subject to this budget constraint. Figure 6 depicts the results of this exercise, assuming the same underlying technologies.

The smaller the elasticity of substitution the greater the reduction in output for any given land restriction. For example, when $\sigma = 0.9$, farmers can still produce 26 percent more output, given the budget constraint substituting nonland inputs for land inputs when land use is restricted to the 1980 level (fig. 6). However, when $\sigma = 0.1$, farmers can produce only 7.8 percent more output. If land is restricted to 70 percent of 1980 levels, production decreases to 75.6 percent of 1980 production when $\sigma = 0.1$, but if $\sigma = 0.9$, production increases 13.7 percent with the same restrictions. The higher the elasticity of substitution the smaller the output effect of a land restriction program. Alternatively, the higher the elasticity of substitution the greater the agricultural output despite acreage constraints.

Figure 6

Output Reductions with Land Restrictions and Substitution



The slope measurements of the output curves for any particular land use in figure 6 take on a particular economic meaning if output prices are fixed (subsidized through a target price system). In this case, the slope values when multiplied by output prices are equal to the incremental value or marginal revenue product of an additional unit of land. Marginal revenue products are greater when σ is smaller, indicating the relatively greater economic importance of an additional unit of land to a farmer faced with limited substitution potential.

Conclusion

With a relatively simple model, I have demonstrated the importance of the substitution concept in the discussion of agricultural capacity. Although there are many econometric and agricultural engineering studies of input substitution, each empirical study has a variety of defects, and no definitive estimate of the elasticity of substitution is available. The weight of evidence suggests that this elasticity lies between 0.3 and 0.7. By presenting the agricultural economic impacts of

alternative land use restrictions in figure 6 as well as the impacts of the assumed input price trajectories for target output levels in figures 2-5, I have illustrated the dramatic differences that result from alternative elasticity measures encompassing this range. The input substitution potential measured by the elasticity of substitution, therefore, is particularly important when one assesses the economic impacts of relative input price changes and land use policies. However, the issue as to value(s) of the elasticity has not been resolved, there is some evidence of slightly higher and lower values than the 0.3-0.7 range. It is essential, therefore, that any improved analysis of agricultural capacity provide careful specification of input substitution potential. Moreover, as the knowledge base increases and more ways of producing a given output become available, there is indeed potential for the elasticity of substitution to grow over time. Higher elasticities of substitution imply greater farmer flexibility in the long run to produce sufficient food at relatively low prices. If such elasticity measures are accurate, the agricultural capacity debate may be less important than it appears.

There are, of course, aggregation and separability problems when one assumes the existence of either cost or dual production functions. This article merely offers a simple abstraction that may help sharpen the agricultural capacity debate in world food outlook analysis. For we often assume that $\sigma = 0$, yet we observe here that relaxing this assumption can dramatically change the results of an economic analysis. We do so because of data limitations and other reasons, but the results can be most damaging to policy analysis. I submit that one of the reasons for the ineffectiveness of land programs designed to deal with crop surpluses is that we typically underestimate the value of σ. Moreover, substitution can go both ways. Although the growth rate of yields of many domestic crops appears to be slowing, this slowdown may be attributed to relative input price changes as land is substituted for nonland inputs and not necessarily to a slowdown in technological change. In countries where the rental price of land and capital are substantially higher than in the United States, it is not uncommon to find higher yields, more fertilizers, and more labor used in crop production. Yet experiences of farmers in Japan, Western Europe, Israel, New Zealand, and other countries contributes to the knowledge base in North America and provides the potential for greater agricultural flexibility in the upcoming decades. Other problems, such as current economic and agricultural policies in both developed and underdeveloped countries, could take precedence in the debate over the ability of the U.S. agricultural sector to supply greater quantities of food at profitable farm prices.

References

(1) Allen, R. *Mathematical Analysis for Economists.* New York: MacMillan and Co., 1960.

(2) Binswanger, H. "The Measurement of Technical Change Biases with Many Factors of Production," *American Economic Review,* Vol. 64, 1974, pp. 964-76.

(3) Fuss, M. A., and D. McFadden. *Production Economics: A Dual Approach to Theory and Applications.* Vol. I and II. New York: North-Holland Publishing Company, 1978.

(4) Huffman, W., and R. Evenson. "U.S. Agricultural Productivity and Public Policy: A Many Input-Many Output System." Unpublished manuscript presented at the Econometrics Society Meetings in New York City, N.Y., Dec. 1982.

(5) Lau, L., and P. Yotopoulos. "Profit, Supply, and Factor Demand Functions," *American Journal of Agricultural Economics,* Vol. 54, 1972, pp. 11-18.

(6) Lopez, R. "Applications of Dual Theory to Agriculture," *Western Journal of Agricultural Economics,* Vol. 7, 1982, pp. 510-19.

(7) Mundlak, Y. "Elasticities of Substitution and the Theory of Derived Demand," *Review of Economics Studies,* Vol. 35, 1968, pp. 225-36.

(8) Ray S. "Translog Cost Analysis of U.S. Agriculture, 1939-77," *American Journal of Agricultural Economics,* Vol. 7, 1982, pp. 490-98.

(9) Shephard, R. *Cost and Production Functions.* Princeton, N.J.: Princeton Univ. Press, 1953.

(10) Spinks, T., and D. Dahl. *Inputs Used in U.S. Farm Production.* BLA-19. U.S. Dept. of Agr., Econ. Res. Serv., 1981.

(11) U.S. Department of Agriculture. *Agricultural Statistics, 1982.* 1983.

(12) ________. *Will There Be Enough Food: 1981 Yearbook of Agriculture.* 1981.

Appendix

I employ a CES cost function which is self dual as an example of a knowledge base parameterization. Self dual simply means that the cost function is associated uniquely with a CES production function of the form:

$$Q = (a_N N^{(\sigma-1)/\sigma} + a_L L^{(\sigma-1)/\sigma})^{\sigma/(\sigma-1)}$$

Let the cost function be defined as:

$$C(P_N, P_L, Q) = Q^{1/r}(a_N P_N^{-b} + a_L P_L^{-b})^{-1/b}$$

where a_N, a_L, r, and b are parameters. In this expression, r denotes the degree of homogeneity of the underlying production function and $b = 1-\sigma$ where σ is the elasticity of subsitution. The optimal input equations for N and L are given:

$$N=Q^{1/r} a_N P_N^{-\sigma} \{(a_N P_N^{-b} + a_L P_L^{-b})^{-1/b}\}^{\sigma}$$

$$L=Q^{1/r} a_L P_L^{-\sigma} \{(a_N P_N^{-b} + a_L P_L^{-b})^{-1/b}\}^{\sigma}$$

Note if $\sigma \to 0$ and $r = 1$, then the demand functions for N and L are simply given as a fixed coefficient Leontief input demand function with no input price sensitivity:

$$N=a_N Q$$
$$L=a_L Q$$

Alternatively, if $\sigma \to 1$, then a_N and a_L take on a new meaning as constant cost minimizing factor shares given by:

$$P_N N/C = a_N$$
$$P_L L/C = a_L$$

Because the benchmark values of C, Q, P_N, P_L, N, and L are set equal to 100 for 1980, it is possible to solve for parameters a_N and a_L in the cost function if we impose constant returns to scale—that is, r = 1. Imposing the trajectories of Q, P_N, and P_L for 1980-2000, it is possible to solve for C, C/Q, Q/L, and L, for each time period for each σ. These results are contained in figures 2-5.

For the results displayed in figure 6, I fix C at 130 and solve for the parameters a_N and a_L in the CES production function where N and L are set initially at levels 30 percent greater than 1980 levels and input prices are held fixed. Once values for a_N and a_L are obtained, I restrict the land use to between 70 percent less and 30 percent more than 1980 land use. Recall L in 1980 = 100. I then solve for Q subject to the constraint that $P_N N+P_L L = 130$.

Wheat Price: Past and Future Levels and Volatility

By Clark Edwards*

When world food markets were burgeoning during the seventies, people became concerned about longrun food shortages and higher real food prices. When the markets collapsed during the early eighties and food surpluses were again forthcoming from U.S. agriculture, people became concerned about longrun excess capacity and the prospect of declining real prices received by farmers. Through the muddle, a third and more reasonable view emerged. Although shortrun changes in the real level of food prices can be relatively large, the longrun pressures either up or down are not great and the changes are too close to call. The best bet is to predict that the real food price will not change in the long run regardless of how volatile it is in the short run or how wide the swings are in the intermediate run.

I wondered what history has to say about these three views. I decided to examine the price history of a single commodity. I arbitrarily chose wheat despite inherent difficulties with using the price received by farmers for wheat as a proxy for consumers' food prices. Wheat products account for a small percentage of total food outlays; they even account for a small percentage of retail outlays for products that include wheat. Given the trend for increased value added to wheat products in the form of transportation, processing, packaging, and other services, the margin is rising between the price received by farmers for wheat and retail prices of wheat products. Therefore, a stable consumer price level is consistent with a decreasing price of wheat. It is unfortunate, for the purposes of this analysis, that there is no retail price of wheat. Nonetheless, wheat is an important staple in the world food supply, and it is a substitute for other foods as well as for feed for livestock. Furthermore, it is the price received by farmers that induces the quantity supplied, not the retail price. General economic phenomena such as wars, depressions, and world food crises are reflected in the price of wheat. This relationship implies that an enduring worldwide scarcity of food will be reflected in a rising wheat price and worldwide abundance will be reflected in a falling price.

Agricultural Statistics: 1983 lists the price of No. 1 Hard Winter wheat, ordinary protein, at Kansas City, as far back as 1968. The 1972 issue shows the series to 1929. *Historical Statistics of the United States: Colonial Times to 1970* takes the series back to 1800. However, the footnotes to the tables warn that the data source changes from time to time. For example, the series reports No. 2 wheat prior to 1961, and there are other changes in market reporting. However, a change of a different nature occurred in 1913. Immediately prior to 1913 the Chicago market was used, and still other markets and other classifications of wheat were used in earlier years. I decided to stop there and use the series as reported for Kansas City from 1913 to the present. (I am telling you this because I think it is an important principle of agricultural economics research that what we study and what we conclude depend a great deal on what data are available). The series is shown in figure 1.

The price of wheat at Kansas City shows the relatively high price of food during and immediately after World War I. The agricultural depression of the twenties is clear as is the further downward pressure on price during the Great Depression of the thirties. The price held close to its World War II high throughout most of the fifties and sixties; a gradual downtrend is apparent through that period. It is also apparent that annual price fluctuations were limited during that period. The fifties and sixties were years of massive Government programs which bolstered the domestic price above the world price and supported farm income. One effect of these programs was to reduce price fluctuation. The downtrend during the fifties and sixties reflects policy adjustments to work off accumulated stocks of wheat that had not cleared the market at the supported price, and it reflects accommodation to the fact that the domestic price

*The author is an economist with the National Economics Division, ERS.

was above the downward-trending world price. Exposure of the domestic price to world trade during the world food crisis of the seventies drove the price of wheat to a historic high and reintroduced wide annual price fluctuations.

General economic phenomena are reflected in this commodity price series, phenomena that also affected prices of other commodites at both the producer and consumer levels. For this reason, I intend to derive some general inferences about the future price of food from this history of the price of wheat at Kansas City.

One gets the sense from figure 1 that the price of food has been rising during the 20th century and that the major swings in price reflect major influences such as wars and depressions. The major swings are real, but the price level rise may be illusory; it is important to know the price of wheat relative to other prices. From the producer's point of view, the price of wheat relative to the cost of production is important. To the farm family, it may be the price of wheat relative to the cost of food, clothing, and shelter. The nonfarm consumer's view is close to that of the farm family: What happened to the price of food relative to other things consumers buy? This comparison suggests deflating the price of wheat with the consumer price index. *Historical Statistics of the United States* supplements current U.S. Department of Labor sources with the consumer price index to 1913. Figure 2 shows this series. World Wars I and II are apparent in the series as is the Great Depression. However, the dramatic portion of the figure is the rapid rise in the cost of living since the midsixties. Deflating the current wheat price in figure 1 with the consumer price index in figure 2 produces the real price of wheat in 1967 dollars, shown in figure 3.

The years of war, depression, and food crisis appear in figure 3 as clearly as in figure 1, as do the periods of relatively high annual price fluctuations before the fifties and after the sixties. What is different is that figure 3 gives the impression of a downtrend in real price whereas figure 1 gives the impression of an uptrend in nominal price. Whether you conclude from figure 3 that the real price of food is trending downward or not depends on which years you pick for the end points. Certainly if you accept the arbitrary beginning point shown in the figure, 1913, the real price decreases over the years. A regression of the real price of wheat on time reveals that the downtrend averages more than 2 cents per bushel per year; the coefficient is significant with a t ratio of 5. However, if you start with the early twenties, the downtrend is not so clear, and if you start with the early thirties, you can almost see an uptrend. Figure 4 shows one way to think about this dilemma.

Figure 4 depicts a 10-year moving average price of wheat. To interpret a moving average, consider an observer during the year 1980. The expected price of wheat for the year 1980 is taken to be the central tendency for the years 1970 to 1979. A year later, 1970 is dropped from the calculation and 1980 is added to form an expectation for 1981. The moving average concept strikes some as fuzzy because a single observation keeps showing up with the same weight in different sample means. For example, the relatively high wheat price of 1973 is in the 1980 sample and is there again in the 1981 sample. It will suddenly be dropped from the 1983 sample. Some researchers prefer, therefore, to show, for example, an average for each decade. Either technique can be used to tell the story. The moving average technique has the advantage of depicting a continuous flow which removes the annual fluctuations and makes the major real price swings related to war, depression, and food crisis more readily discernible. It also gives the clear impression that the peak real price following World War II was below the World War I peak and that the price of wheat during the seventies was below the depressed price of the thirties. This way of thinking about the real price of wheat clearly suggests a longrun downtrend.

What about price volatility? Inspection of nominal price in figure 1 suggests that the price of wheat was relatively stable during the fifties and sixties and was relatively volatile before and after. Inspection of real price in figure 3 suggests the same conclusion. Annual volatility, of course, is removed in the 10-year moving averages in figure 4. The standard deviation is a useful measure of dispersion. A range of plus and minus one standard deviation around a central value captures about two-thirds of the observations. Figure 5 shows the 10-year moving standard deviation for the nominal wheat price.

To see how figure 5 is interpreted, consider that the standard deviation was about 50 cents per bushel for the decade that ended in 1950.

This means that about 7 of the previous 10 prices for wheat were within (plus or minus) 50 cents of the 1950 price. Figure 5 shows the variation in wheat price was relatively small from the mid-fifties through the early seventies. Figure 6 shows the 10-year moving standard deviation for the real wheat price. It also suggests relatively stable prices through the fifties and sixties. The major difference in the interpretation of figure 5 relative to figure 6 is that the nominal price series suggests a very large increase in volatility during the seventies, whereas the real price series shows a rise that may be called moderate in comparison with the volatility associated with the post-World War I period.

The coefficient of variation is the ratio of the standard deviation to the mean. The advantage to using the coefficient of variation instead of the standard deviation as an indicator of dispersion is that because the unit of measure (dollars per bushel in this case) is in both the numerator and denominator, it cancels out, and a relative measure of dispersion is achieved which is independent of the unit of measure. This property means that the measure is invariant with respect to whether quantity is measured in bushels or tons and whether price is measured in dollars or yen. And it raises the question as to whether the general price level (inflation) is also removed.

The coefficient of variation for the nominal price is shown in figure 7 and for the real price in figure 8. Both figures show what was already clear from figure 1—that the wheat price was more stable during the fifties and sixties than before or since. Figures 7 and 8 each tell about the same story with respect to the degree of volatility before the fifties and after the seventies. The question raised by comparing standard deviations of the nominal and real series is resolved. We do not need to decide whether or not the post-World War I period was more volatile than the seventies; figures 7 and 8 suggest that the relative degree of volatility was about the same. Inasmuch as the coefficients of variation for the nominal and real prices tell approximately (but not exactly) the same story, whereas the standard deviations for nominal and real prices tell different stories, one can infer that the coefficient of variation for the nominal series approximately (but not exactly) removes the effect of inflation.

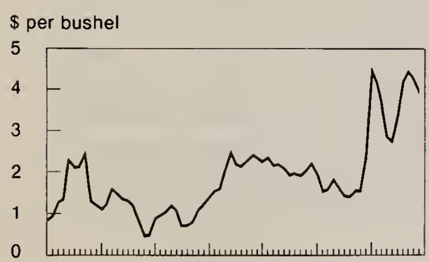
Figure 9 summarizes everything I have said about the price of wheat. However, figure 9 is a fairly abstract way of presenting information about the actual series shown in figure 1. Let's assume for the sake of argument that the series in figure 1 represents the real world which we seek to describe and that we know concretely what the data in figure 1 represent. I deflated that series by the index number known as the consumer price index and then calculated a 10-year moving average of the real wheat price. I also calculated a 10-year moving standard deviation. Consider, for each year since 1923, a range of wheat price from one standard deviation below to one standard deviation above the 10-year average. Now, like the Cheshire cat, let things start to vanish—the nominal price of wheat, the real price, and the moving average—until nothing is left but the end points of the range. It is the remaining smile that is depicted in figure 9.

Figure 9 indicates the longrun downtrend in the real price of wheat; the major swings related to war, depression, and food crisis; and the degree of annual volatility around the expected price. One can see that the range of annual fluctuation was relatively narrow during the fifties and sixties. During the seventies, the degree of shortrun price volatility appears to have returned to its earlier character.

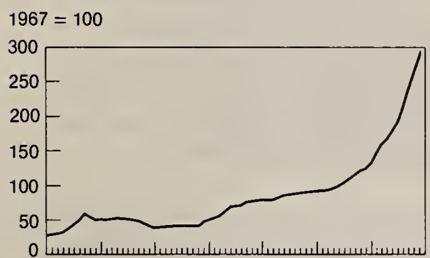
Several views of future food prices have been aired in the literature. The history I have reviewed here of one major food commodity in one major market over most of this century suggests a longrun downtrend and a relatively high degree of volatility. If the price of wheat at Kansas City is a useful proxy for food prices, then those who predict increasing real food prices in coming decades, who suggest that the best bet is to predict that real food prices will not change, or who anticipates a return to the relative price stability of the fifties and sixties are really calling for a fundamental change in the longrun trend.

A Graphic History of the Price of Wheat: 1913-83

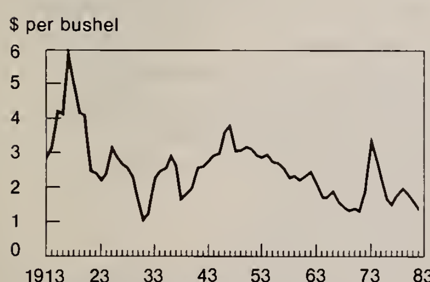
1. **Wheat Price: Current Dollars**



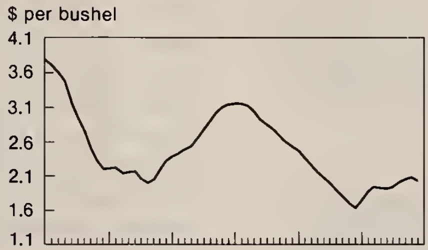
2. **Consumer Price Index**



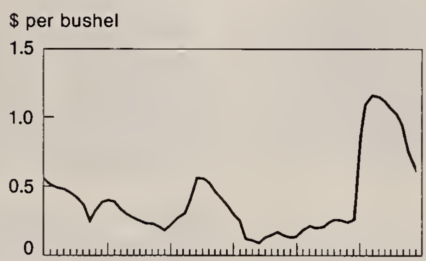
3. **Real Wheat Price: 1967 Dollars**



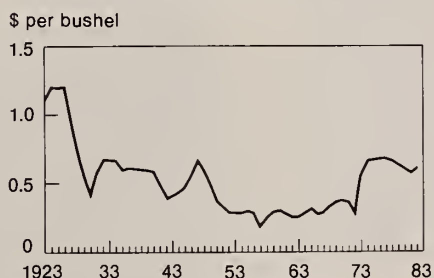
4. **Real Wheat Price: 10-Year Moving Average, 1967 Dollars**



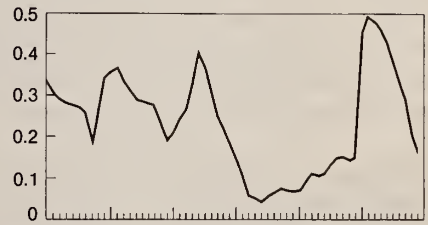
5. **Wheat Price: 10-Year Moving Standard Deviation**



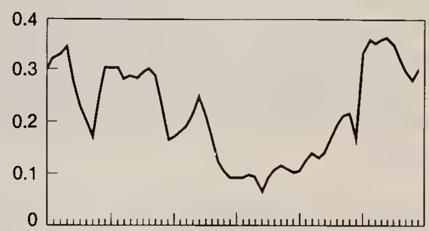
6. **Real Wheat Price: 10-Year Moving Standard Deviation, 1967 Dollars**



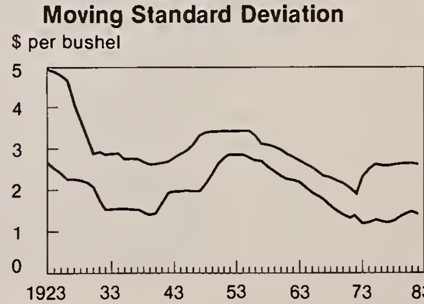
7. **Wheat Price: 10-Year Moving Coefficient of Variation**



8. **Real Wheat Price: 10-Year Moving Coefficient of Variation, 1967 Dollars**



9. **Moving Average Real Wheat Price: Plus and Minus One Moving Standard Deviation**



The Federal Lands Revisited

Marion Clawson. Washington, D.C.: Resources for the Future (distributed by the Johns Hopkins University Press, Baltimore and London), 1983, 302 pp., $25.00 (cloth), $8.95 (paper).

Reviewed by Robert F. Boxley*

At the beginning of the current administration, much was made of the Sagebrush Rebellion and the drive for making public lands private. As an observer with at least a passing interest in the issue, I recall my frustrations with the sketchy documentation of the proposals by those arguing for privatization and with the tendency of the debate to be cast in absolute all or nothing terms.

Although the Sagebrush Rebellion has since been quelled, Marion Clawson's new book, *The Federal Lands Revisited*, provides a lucid commentary on both the battle past and the war ahead. Clawson states that, some 20 years from now, the late seventies and eighties may appear as an important juncture in the evolving Federal land history. He believes that now is a propitious time to reexamine basic Federal land policy, and he argues: "It is wholly possible to invent new institutions and new arrangements for the use of the federal lands" (p. xvi).

In three chapters central to this argument, Clawson outlines how changes might be accomplished. He presents the retentionist's case for continued Federal landownership, the disposer's case for privatization, and the political economist's case for new institutions and arrangements. As enumerated by Clawson, the middle ground is broad. Options include retention of current public lands with greatly improved management; transfer to the States; disposal to private ownership; management by public or mixed public-private corporations; and large-scale, long-term leasing. The long-term lease alternative receives the most attention from Clawson.

Clawson also proposes an innovative "pullback" procedure. Under the pullback concept, individuals or groups could apply for a tract of Federal land for any use they choose, but any other person or group would have a limited time between filing an initial application and granting of the lease or making the sale in which to pull back a part of the area applied for. Clawson sees the pullback provision as a device for introducing competition among potential users of Federal lands and for promoting bargaining among competitors. He argues the pullback provision would reduce collusion; guarantee adherence to bargains, once established; reduce incentives to use delaying tactics; and provide a better mechanism for negotiating among rival private interest groups.

A not incidental service Clawson provides in this section of the book is his careful documentation of the rather sparse privatization literature. The case for privatization was made principally in speeches and in trade publications rather than in professional journals and books. Clawson has done a good job of documentation throughout the book, especially in his discussion of privatization.

Readers will get far more than blueprints for new institutions and new arrangements for using Federal lands. They will also find a concise minihistory of Federal lands; a comprehensive overview of current Federal land use, planning, and management issues; a discussion of the special problems of intermingled Federal-private landownership; and an analysis of the difficulties of achieving public participation in public land-management decisions. Readers will even get what Clawson ruefully concedes is *de rigueur* in books of this nature—a chapter on the need for further research.

I assume that most readers of this journal are already familiar with the prolific writings of Marion Clawson. For 45 years he has been professionally concerned with the Federal lands of the United States: as an economist in the Bureau of Agricultural Economics of the U.S. Department of Agriculture, as regional administrator and director of the Bureau of Land Management in the U.S. Department of the Interior, and as a member of the research staff of Resources for the Future (RFF). Of his experiences he says, "I scarcely could fail to have learned something about these lands. In

*The reviewer is an agricultural economist with the Natural Resource Economics Division, ERS.

fact, I have acquired a great deal of knowledge, perhaps a little wisdom, and surely my fair share of biases and prejudices." He notes that the book is more personal in tone than the usual research volume from RFF. It is, and because it is, it is also a delight to read.

In Earlier Issues

In using statistical procedures in the analysis of prices the student must keep constantly in mind that the numerical or graphic results, no matter how good they may be, tell nothing about the reasons for the relationships. These reasons must be found in the general knowledge of the relationships and the general logic of the situation. . . .

Warren C. Waite and Harry C. Trelogan
Vol. 1, No. 1, Jan. 1949

. . . advertising may substantially affect national food choice. By raising prices on heavily advertised products, many consumers are forced to substitute less desirable brands in the same product category. Advertising probably shifts interindustry demand as well as interbrand demand in the long run. Advertising may be partially responsible for the notable shift in preference away from milk, fruit juices, and water (which are less advertised) to artificially fruit-flavored drinks, soft drinks, tea, and alcoholic beverages (all of which are heavily advertised).

John M. Connor
Vol. 33, No. 1, Jan. 1981

*U.S. GOVERNMENT PRINTING OFFICE : 1984 0-420-932/FS-2204

Suggestions for Submitting Manuscripts for *Agricultural Economics Research*

Contributors can expedite reviewing and printing of their articles by doing these things:

1. SOURCE. Indicate in a memorandum how the material submitted is related to the economic research program of the U.S. Department of Agriculture and its cooperating agencies. State your own connection with the program.
2. CLEARANCE. Obtain any approval required in your own agency or institution before sending your manuscript to one of the editors of *Agricultural Economics Research.* Attach a copy of such approval to the manuscript.
3. ORIGINALITY OF MATERIAL. It is our policy to print original material. We consider alternative treatments of material published elsewhere, but such treatments need to differ substantially from the original approach. When submitting your manuscript, identify, if applicable, related material either published or submitted for publication.
4. ABSTRACT. Include an abstract and at least three keywords when you submit your article. The abstract should not exceed 100 words. It should report the major findings of your research.
5. NUMBER OF COPIES. Submit three good copies.
6. TYPING. Double space everything, including abstract, footnotes, and references.
7. FOOTNOTES. Number consecutively throughout the article.
8. REFERENCES. Check all references carefully for accuracy and completeness.
9. CHARTS AND OTHER ARTWORK. Use charts sparingly. Keep design as simple as possible. Submit all artwork in draft rather than in final form, accompanied by neatly prepared pages with essential data for replotting.

Researchers interested in a bibliography arranged alphabetically by author of materials in *Agricultural Economics Research* from the first issue in 1949 through the current issue should write Gerald Schluter, Economist, ERS/NED, U.S. Department of Agriculture, Washington, D.C. 20250.

Readers interested in a chronological bibliography of materials in *Agricultural Economics Research* from the first issue in 1949 through the last issue in 1979 should write Gerald Schluter, Economist, ERS/NED, U.S. Department of Agriculture, Washington, D.C. 20250.

United States
Department of Agriculture

Washington, D.C.
20250

OFFICIAL BUSINESS
Penalty for Private Use, $300

Postage and Fees Paid
U.S. Department of Agriculture
AGR-101

U.S. MAIL

THIRD CLASS BULK RATE

```
* AGER HOLE 453E ISSDUE001R     1 *
* ERLING HOLE                     *
* 6453 OAKWOOD DR                 *
* FALLS CHURCH   VA   22041       *
*                                 *
```

3